

PORPHYRY AU PAYS DES PAESTANS
Usages d'un outil d'analyse qualitative de documents
par des étudiantes de maîtrise en iconographie grecque

Aurélien BÉNEL
Laboratoire Tech-CICO (Institut Charles Delaunay)
Université de Technologie de Troyes

SOMMAIRE

1. Introduction
2. Corpus
3. Source
4. Point de vue
5. Portfolio
6. Conclusion

Résumé : Cette communication livre un retour d'expérience sur l'introduction de *Porphyry*, logiciel d'annotation sociale, dans l'activité d'analyse d'un corpus de photographies par une équipe d'historiens d'art. Leur corpus comprend les photographies (profil, détail du recto, détail du verso) des 300 vases de la région de Paestum (en Campanie) représentant des scènes de banquet ou liées à Dionysos. Au printemps 2005, trois étudiantes ont contribué à l'étude iconographique de ces vases, dans le cadre de leur maîtrise en sciences de l'antiquité.

Le compte-rendu de cette expérience par le responsable scientifique de l'équipe nous intéresse à plus d'un titre : par la description de l'activité instrumentée d'annotation dans un « jargon » différent de celui des concepteurs du logiciel, par l'absence de mention à des fonctionnalités jusqu'alors considérées comme capitales, enfin, par des critiques, en apparence anodines, qui pourraient remettre en question certaines positions théoriques.

Par exemple, pourquoi regretter le fait que l'on ne puisse pas « renommer un descripteur » ? Sans être pour autant un concept, serait-il tout de même plus qu'une chaîne de caractères ? Le mot « descripteur » est-il alors bien choisi ? De même, pourquoi dénoncer l'absence de comptage automatique dans un outil qui se veut avant tout qualitatif ? D'ailleurs, que s'agit-t-il de compter : les fragments, les sources photographiques, ou bien, indirectement, les vases ? N'y a-t-il pas alors un surprenant « glissement » référentiel ? Tel est le genre de questions qui seront abordées dans cette communication.

Remerciements : Cet article constitue le premier retour d'expérience d'une collaboration de trois ans entre le réseau ARTCADHi-CNRS¹ et le laboratoire CRATA². Cette collaboration n'a pu être possible que grâce au travail et à la persévérance des animateurs scientifiques et de leurs équipes. Nous remercions particulièrement Andrea Iacovella, Jean-Marc Luce, Véronique Pouyadou, Samuel Gesche, Pascale Jacquet et ses étudiantes.

1. Introduction

Comme aurait pu le dire *Qohéleth*, il y a un moment pour tout et un temps pour chaque chose dans la recherche en informatique : un temps pour lire et un temps pour écrire, un temps pour théoriser et un temps pour développer des logiciels, un temps pour former les usagers et un temps pour les observer travailler, un temps pour mettre à l'épreuve les hypothèses et un temps pour en forger de nouvelles.

Dans cette communication, nous allons donner un retour d'expérience sur l'introduction de *Porphyry*, logiciel d'annotation sociale (Iacovella et al., 2005 ; Bénel, 2003), dans l'activité d'analyse d'un corpus de photographies par une équipe d'historiens d'art. Le corpus comprend les photographies (profil, détail du recto, détail du verso) des 300 vases de la région de Paestum (en Campanie) représentant des scènes de banquet ou liées à Dionysos (Pouyadou, 2001). Au

¹ Réseau ARTCADHi-CNRS : « Atelier de recherches transdisciplinaires sur la construction du sens en archéologie et autres disciplines historiques », <<http://www.porphyry.org/>>.

² Laboratoire CRATA : « Culture, représentations, archéologie, théâtre antique », <<http://www.univ-tlse2.fr/crata/>>.

printemps 2005, trois étudiantes ont contribué à l'étude iconographique de ces vases, dans le cadre de leur maîtrise en sciences de l'antiquité.

Le compte-rendu de cette expérience par le responsable scientifique de l'équipe nous intéresse à plus d'un titre. Par exemple, l'usage du nom des primitives de notre outil (*descripteurs* : *facettes*, *descripteurs ordinaires*, *identifiants*), usage souvent difficile, parfois « fautif », nous amènera à nous questionner sur notre « jargon » et à le simplifier. De même, l'absence de mention à des fonctionnalités jusqu'alors considérées comme capitales pourra nous amener à nous interroger sur ce qui les rend inopérantes. L'analyse de critiques, en apparence purement techniques, pourra également remettre en question certaines positions théoriques.

Pour analyser ce compte-rendu d'usage, nous prendrons pour grille les quatre objets principaux autour desquels l'activité instrumentée se déploie : le corpus, la source, le point de vue et le portfolio.

2. Corpus

Les photographies de chaque vase ayant été numérisées, l'expert les importe dans le système pour qu'elles deviennent des *sources*. Survient alors un premier problème signalé dans le compte-rendu :

Inconvénients : [...] 2 - L'impossibilité d'introduire dans le nom d'un descripteur un mot qui figure dans le nom de la facette. (Luce, 2005)

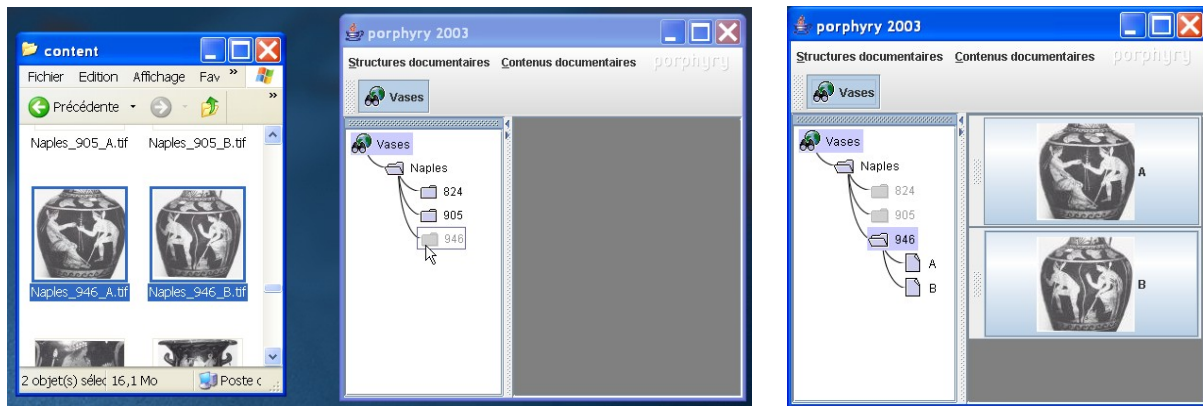


Fig. 1 : Séquence d'importation de sources (par un glisser-déposer) lors de la constitution du corpus.

Au moment de l'import, un *identifiant d'objet documentaire* est assigné à la source. 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 aléatoires¹. 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 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 ». La création de ce réseau, tâche répétitive et sans grand intérêt pour l'expert, est censée être rendue « transparente » grâce à un algorithme simple sur les chaînes de caractères des noms de fichiers à importer, à en croire la doléance mentionnée plus haut, elle n'en est que plus obscure...

Si nous avons choisi de discuter de ce problème, en apparence trivial, 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

¹ L'identifiant d'objet documentaire reste d'ailleurs, aujourd'hui encore, caché de l'utilisateur.

interpréter, nous éviterait les pièges de tous les intégrismes (le positivisme 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 ?

3. Source

Une fois le corpus établi, même partiellement, le travail sur les sources peut commencer. Le responsable souligne :

Avantages : [...] 8 - Système moins contraignant que le travail sur des tableaux. Le système maintient l'opérateur en constant contact avec les images. Les étudiantes ont souligné le caractère ludique des manipulations. (Luce, 2005)

Parmi les manipulations permises par le système sur les sources, une seule laisse des traces : celle consistant à sélectionner des fragments. Or, il n'en est fait aucune mention dans le compte-rendu. L'observation du corpus nous révèle d'ailleurs 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.

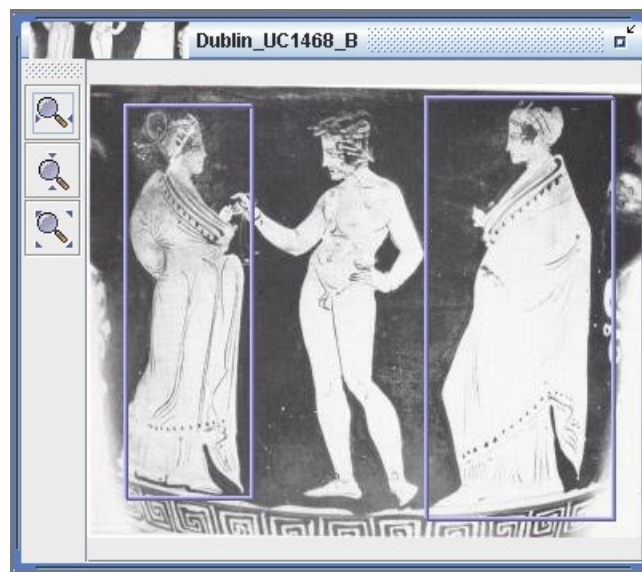


Fig. 2 : Cas fictif de sélection dans une source de fragments répondant à la même thématique.

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 curiosité due au 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 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.

4. Point de vue

Les ensembles de fragments (ou de sources) vont ensuite être décrits par l'expert, à l'aide de *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.

Mais si le descripteur est un terme, comment interpréter le commentaire suivant ?

Inconvénients : 1 - L'impossibilité de renommer un descripteur. C'est très ennuyeux en cas d'erreur. On est obligé de refaire le descripteur. (Luce, 2005)

Plutôt que par son nom, le « descripteur » ne serait-il pas davantage caractérisé, comme le laisse entendre le commentaire suivant, par les fragments et les sources qu'il contient déjà ?

Avantages : [...] 2 - Possibilité de faire des liens avec l'apparition immédiate des images. (Luce, 2005)

Tout porte à croire que nous sommes en présence d'une *collection* (Rousseaux, 2006), 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.

Une fois créées, les collections peuvent être organisées entre elles :

Avantages : 1 - La possibilité de créer des rubriques à l'infini pour réaliser un relevé iconographique précis. Cette possibilité est très appréciable par rapport aux bases de données. (Luce, 2005)

Inconvénients : [...] 5 - Quand les liens sont nombreux, l'arborescence devient vite illisible. (Luce, 2005)

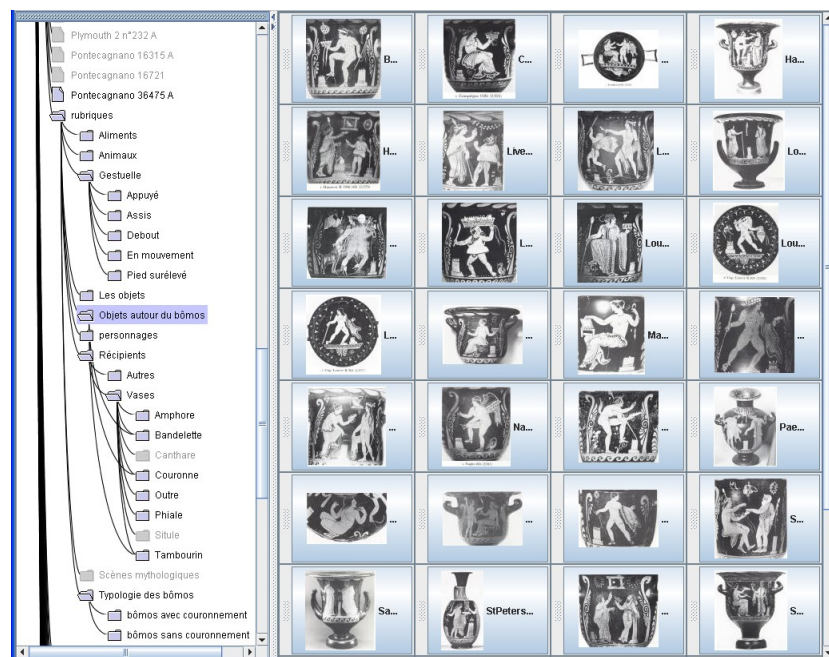


Fig. 3 : Point de vue portant sur l'étude des scènes où figurent des autels.

Le réseau des collections comporte les inconvénients et les avantages de sa complexité. Polyhiérarchique, il permet à un même objet documentaire d'être décrit selon plusieurs dimensions

d'analyse et même à une collection d'être à l'intersection de plusieurs collections (ce qui est particulièrement utile pour le temps, l'espace, ou encore les rôles d'un objet). Une telle structure l'empêche par contre d'être affiché dans son ensemble sans que ses arcs ne se coupent. Une visualisation interactive pourrait être mise en place de sorte que seuls certains arcs soient affichés en réponse aux actions de l'utilisateur.

5. Portfolio

Corpus et points de vue sont rassemblés dans un portfolio. C'est le moment de feuilleter ce corpus enrichi, le moment de découvrir le corpus sous un jour nouveau.

Avantages : [...] 4 - Créer des associations auxquelles on n'aurait pas pensé ; 5 - Le logiciel est particulièrement intéressant pour mettre en évidence des cas exceptionnels. (Luce, 2005)

Les commentaires que nous venons de citer font référence à la possibilité de sélectionner plusieurs collections et de voir en retour, d'une part, les objets documentaires présents à leur intersection, d'autre part, le statut de co-occurrence des autres collections (disjonction, conjonction partielle, conjonction totale).

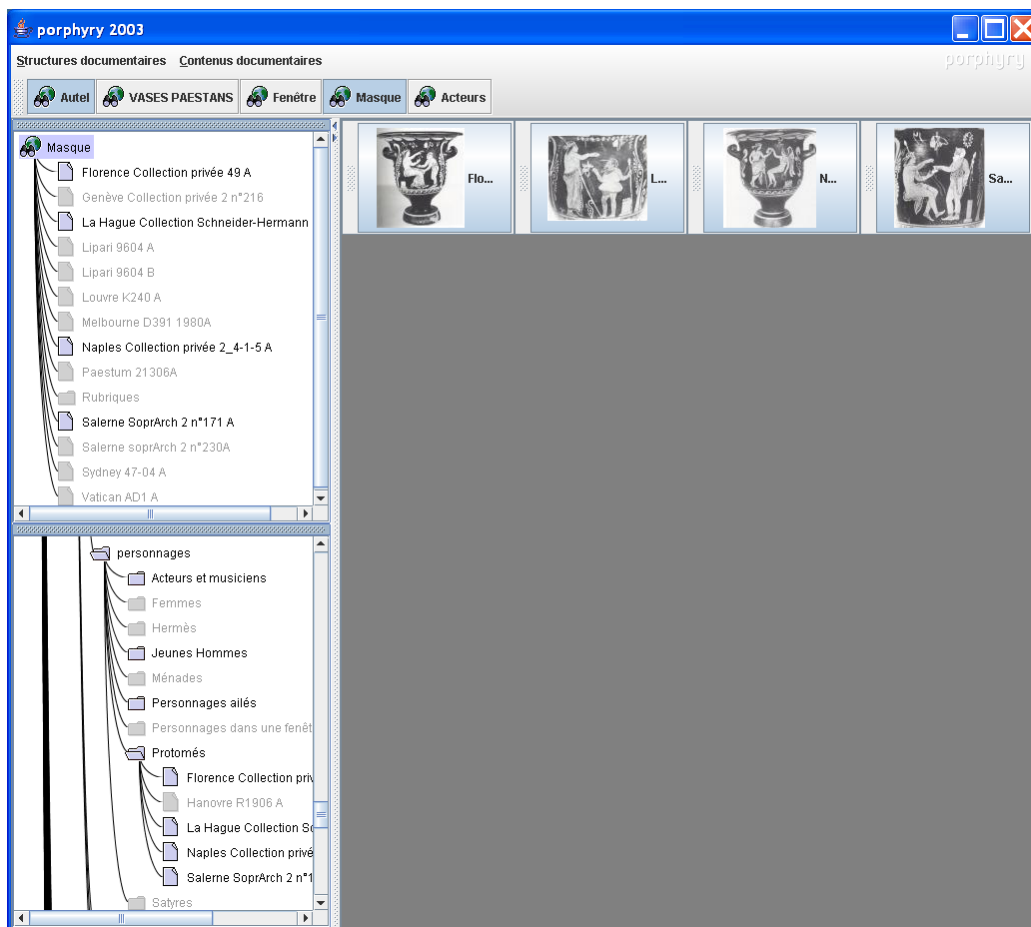


Fig. 4 : Statut de co-occurrence entre les masques et les différents types de personnage autour de l'autel.

Inconvénients : [...] 4 - On ne peut pas demander à l'opérateur de constamment compter les vignettes qui s'affichent. Il faut un outil qui les compte automatiquement. (Luce, 2005)

Pourquoi dénoncer l'absence de comptage automatique dans 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, est 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 actuelle n'affiche pas l'ensemble des objets documentaires mais juste

quelques exemples prototypiques. Les usagers ont cependant réussi à contourner ce comportement du logiciel en dupliquant à la racine du point de vue chaque objet documentaire. Par ailleurs, que s'agit-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 (Iacovella *et al.*, 2005).

Les commentaires suivants mettent l'accent sur le caractère social du portfolio, lieu de publication (au sens de « rendre public ») et de confrontation des points de vue, lieu d'accord et de désaccord, de sédimentation et de révolution dans les savoirs.

Avantages : [...] 3 - la possibilité à l'avenir de rentrer plusieurs travaux, chacun peut ensuite bénéficier du travail des autres ; [...] 6 - à partir d'un même corpus de base, possibilité de créer plusieurs facettes sur des thèmes différents, 7 - possibilité de développer des réflexions différentes sur un même corpus. (Luce, 2005)

Au fur et à mesure que de nouveaux groupes travailleront sur porphyre, les corpus vont s'élargir, les liens vont devenir plus nombreux et il sera possible de comparer des points de vue en utilisant les diverses facettes. (Luce, 2005)

Inconvénients : [...] 6 - L'impossibilité d'imprimer autrement que par la capture d'écran qui ne permet pas d'imprimer tout ce qui est sélectionné si le nombre de descripteurs ou des documents ne tient pas dans l'écran, 7 - l'impossibilité de sortir du bureau pour utiliser porphyre, même pour montrer les résultats aux collègues. (Luce, 2005)

Le dernier point évoqué pourrait surprendre. Pourquoi vouloir à tout prix imprimer les résultats et les « sortir du bureau », quand on utilise un environnement visant une continuité numérique entre la lecture et l'écriture ?

On pourrait tout d'abord objecter que « sortir du bureau pour utiliser porphyre » ne sera bientôt plus un problème, les réseaux sans fil se généralisant. Mais le problème est peut-être plus profond : la publication ne nécessiterait-elle pas un changement de médium ? Si l'on peut envisager des « passerelles » vers le Web et le monde du papier, comment le faire sans perdre les propriétés du portfolio numérique ? À moins que ce soit justement le caractère plus stable et moins polémique qui soit recherché dans ces média. Ne risque-t-on pas alors de passer à côté de la révolution d'une herméneutique numérique ?

6. Conclusion

Le parcours que nous venons de tracer à travers un compte-rendu de l'usage de *Porphyry*, nous a permis de poser un certain nombre de questions portant sur les conditions de construction du sens, à plusieurs, à partir de documents. Cette communication est l'occasion rêvée de poser ces questions à d'autres et d'obtenir en retour quelques éléments de réponse. Nous ne doutons pas que ces éléments de réponse configureront ce que seront dans le futur ce type d'outils d'analyse qualitative de documents et plus particulièrement leurs interfaces homme-machine.

La prochaine étape souhaitée serait l'appropriation de l'outil dans un usage quotidien.

Elles [les étudiantes] ont conclu en exprimant leur frustration sur un point : leur mémoire de maîtrise a été faite à l'aide de tableaux. Maintenant qu'elles ont goûté à Porphyre, elles souhaitent pouvoir s'en servir pour leurs propres mémoires. Elles le feront pour leur mémoire de master 2. (Luce, 2005)

Espérons que l'appellation de l'outil par un nom (« Porphyre » et même « porphyre ») différent de son nom officiel (« Porphyry ») soit le gage d'une appropriation déjà en marche !

BIBLIOGRAPHIE

BÉNEL, A. 2003. *Consultation assistée par ordinateur de la documentation en Sciences Humaines : Considérations épistémologiques, solutions opératoires et applications à l'archéologie*, Thèse de doctorat en informatique, INSA de Lyon, décembre 2003. In *Texto ! mars 2004*. Disponible sur : <<http://www.revue-texto.net/Inedits/Benel/Benel.html>>.

IACOVELLA, A., BÉNEL, A., CALABRETTO, S., HELLY, B. 2005. Assistance à l'interprétation dans les bibliothèques numériques pour les sciences historiques, in J.-L. Lebrave (éd.), *La société de l'information et ses enjeux, Actes du colloque de bilan du programme interdisciplinaire « Société de l'information »*, pp. 167-179.

Disponible sur : http://www.porphiry.org/Members/abenel/benel_PSI_05.pdf

LORTAL, G., LEWKOWICZ, M., TODIRACU-COURTIER, A. 2005. Modélisation de l'activité d'annotation discursive pour la conception d'un collecticiel support à l'herméneutique, in *Actes des 16ème journées francophones d'ingénierie des connaissances*, Grenoble, PUG, pp.169-180.

LUCE, J.-M. 2005. *Porphyre à Toulouse*, courriel adressé à A. Iacovella et A. Bénel, 29 juin 2005.

POUYADOU, V. 2001. Dionysos barbu : le sens du poil, *Pallas* 57, pp. 169-183.

ROUSSEAU, F. 2006. La collection, un lieu privilégié pour penser ensemble singularité et synthèse, *EspacesTemps.net*, février 2006.

Disponible sur : <http://espacestems.net/document1836.html>