

Bibliothèques virtuelles humanistes
Centre d'études supérieures de la Renaissance
Université de Tours
59 rue Néricault-Destouches
BP 11238
37013 Tours Cedex 1

**Institut de recherche sur les
Archéomatériaux**
CNRS-Université d'Orléans
Centre Ernest-Babelon
3D rue de la Férolierie
45071 Orléans cedex 2

Appel à communications

Gestion informatisée des écritures anciennes : état des lieux et perspectives

Orléans, 21-22 mai 2013

Le développement des *humanités numériques* offre, depuis deux décennies, des perspectives stimulantes tant pour les travaux des spécialistes que pour la diffusion de leurs résultats auprès du grand public. Les projets impliquant le traitement informatisé de corpus de documents textuels ou iconographiques se multiplient. L'exploitation des données et métadonnées ainsi produites constitue un enjeu important dont témoigne le développement d'outils et de formats d'indexation toujours plus performants.

En amont des problèmes posés par l'indexation des données, un constat s'impose pourtant aux chercheurs : celui de l'inadéquation de la typographie numérique avec les formes graphiques présentes dans les documents anciens. Jeune technologie, l'informatique ne permet pas encore la mise en œuvre de pratiques de transcription satisfaisantes pour les textes anciens. Avant même leur indexation, les données numériques ne sont pas nécessairement codées de manière cohérente.

Si le standard Unicode entend prendre en charge l'ensemble des langues écrites, tant anciennes que modernes, les chercheurs constatent bien souvent son inadéquation aux besoins concrets de leurs travaux. Définissant le caractère comme pure entité linguistique, Unicode exclut d'emblée un grand nombre de variantes graphiques, de ligatures ou de signes abrégatifs dont la présence dans les documents originaux peut révéler une provenance, un état de langue, une pratique dont il importe de garder trace. Plusieurs domaines de recherche (tels l'étude des pratiques orthographiques, l'indexation des légendes monétaires, la transcription de documents épigraphiques, l'analyse des matériels typographiques anciens, le catalogue de manuscrits et d'imprimés...) nécessitent l'utilisation de caractères actuellement absents des codages numériques. Si le travail accompli par le consortium MUFI (*Medieval Unicode Font Initiative*) a permis de combler quelques-unes des lacunes dans le domaine spécifique des manuscrits médiévaux, de nombreux signes doivent encore être identifiés, codés et dessinés :

– *L'identification* des caractères nécessaires au traitement informatique de chaque objet d'étude est rarement menée à bien. Il importe pourtant de répertorier les caractères manquants en leur donnant une définition stricte, et en distinguant les glyphes signifiants des simples variantes graphiques.

– *Le codage* des caractères manquants doit être universel. Pour assurer la pérennité des données numériques produites, il importe donc de veiller à l’harmonisation des pratiques de codage.

– *Le dessin* des caractères, c’est-à-dire la réalisation de polices numériques permettant l’affichage et l’impression des caractères nouvellement codés ne constitue que la partie la plus visible de l’iceberg ; elle n’en est pas pour autant une étape négligeable.

En dépit de vingt années de développement des *humanités numériques*, et malgré les avancées décisives du consortium MUFI, le problème du codage des écritures anciennes demeure donc entier. C’est pourquoi le Centre d’Études supérieures de la Renaissance (CESR, Tours) et l’Institut de Recherche sur les Archéomatériaux (IRAMAT, Orléans) organisent les 21 et 22 mai 2013 à Orléans deux journées d’études consacrées à ces problématiques et destinées tant aux chercheurs issus des lettres, des sciences humaines et sociales, des bibliothèques ou des musées, qu’à ceux exerçant dans le domaine du graphisme ou de l’informatique.

Thématiques

Différentes sessions seront programmées. La liste définitive sera déterminée en fonction des propositions retenues. Nous donnons ci-dessous quelques pistes (cette liste n’étant évidemment pas limitative) :

– Place des caractères anciens dans la représentation de la connaissance.

Réflexion sur la valeur sémantique des glyphes employés dans les documents anciens : quelles informations peuvent-ils contenir ? Comment extraire et représenter cette information dans des systèmes d’indexation sémantique ?

– État des lieux (1). Codages, polices : quelles technologies informatiques pour les écritures anciennes ?

Présentation du fonctionnement et des limites des formats de codage et de police actuels.

– État des lieux (2). Programmes de recherches, réalisations, projets.

Présentation de programmes achevés ou en cours impliquant la gestion informatique d’écritures anciennes.

– Codage ou balisage ?

Où situer la frontière entre ce qui relève du codage et ce qui relève du balisage des textes dans la transmission de l’information ?

– Que transcrire ? Comment transcrire ? Diversité des supports, variations graphiques et états de langues.

Réflexions sur la définition de normes de transcription cohérentes en fonction des besoins propres à chaque projet. Qu’est-ce qu’une version « diplomatique » d’un texte à l’âge numérique ?

– Besoins en caractères.

Projets d’inventaires des caractères manquants pour l’étude ou le catalogage d’objets particuliers (manuscrits, imprimés, sources épigraphiques, numismatiques, etc.).

– **Après le code, la fonte : polices existantes et polices manquantes.**

État des lieux des « besoins en fonte ». Présentation de polices réalisées ou en cours de réalisation, et proposition de modèles pour la gestion et le codage des variantes graphiques.

Les communications, en français ou en anglais, ne devront pas excéder 30 minutes. Les actes des journées d'études feront l'objet d'une publication électronique en libre accès sur internet. Les frais de bouche des intervenants pourront être pris en charge ; les frais de transport et d'hébergement demeureront à la charge des intervenants ou de leurs laboratoires.

Les propositions de communication pourront être adressées **avant le 10 janvier 2013** par voie électronique à Rémi Jimenes (remi.jimenes@univ-tours.fr) et Guillaume Sarah (guillaume.sarah@cnrs-orleans.fr). Elles comprendront un titre provisoire ainsi qu'un résumé d'une demi-page.

Journées d'études organisées par Christine Bénévent (CESR), Rémi Jimenes (CESR) et Guillaume Sarah (IRAMAT), dans le cadre de l'appel à projet MSH Val-de-Loire 2012.

Comité scientifique

- Jacques ANDRÉ (Directeur de recherches retraité, INRIA, Rennes)
- Christine BÉNEVENT (Maître de conférences, CESR, Tours)
- Marc BOMPAIRE (Directeur de recherches, IRAMAT, Orléans)
- Vincent DEBIAIS (Chargé de recherche, CESCUM, Poitiers)
- Marie-Luce DEMONET (Professeur, CESR, Tours ; directrice de la MSH Val-de-Loire)
- Yannis HARALAMBOUS (Directeur d'études, Télécom Bretagne, Brest)
- Rémi JIMENES (Ingénieur d'études, doctorant, CESR, Tours)
- Jean-Yves RAMEL (Professeur, Laboratoire informatique de l'Université de Tours)
- Guillaume SARAH (Chargé de Recherche, IRAMAT, Orléans)
- Marc SMITH (Professeur, École nationale des Chartes, Paris)
- Cécile TREFFORT (Professeur, CESCUM, Poitiers)
- Toshinori UETANI (Ingénieur de Recherches, CESR, Tours)

Merci de nous aider à diffuser largement cet appel en le faisant suivre autour de vous.

