

Codage et classification non supervisée d'un corpus maya : extraire des contextes pour situer l'inconnu par rapport au connu

Mohamed Hallab^{*,**} Bruno Delprat^{***,****} Alain Lelu^{#,##}

^{*}CNAM, La Défense, 15 Ave d'Alsace, 92400 Courbevoie
mohamed.hallab@yahoo.fr

^{**}Université du 7 Novembre, École Supérieure de Technologie et d'Informatique
4, rue des Entrepreneurs Chargaia II - 2035 Tunis-Carthage

^{***}INALCO, École doctorale - 49bis, avenue de la Belle Gabrielle, 75012 Paris
brunodelprat@club-internet.fr

^{****}CÉLIA-CNRS - 7, rue Guy Môquet, 94801 Villejuif cedex

[#]Université de Franche-Comté, Laseldi, 25030 Besançon Cedex
alain.lelu@univ-fcomte.fr

^{##}LORIA, Éq. KIWI, Campus Scientifique - BP 239 - 54506 Vandœuvre-lès-Nancy Cedex

Résumé. L'écriture logossyllabique des anciens Mayas comprend plus de 500 signes et est en bonne partie déchiffrée, avec des degrés de certitude divers. Nous avons appliqué au codex de Dresde, l'un des trois seuls manuscrits qui nous soient parvenus, codé sous \LaTeX avec le système \mayaTeX , notre méthode de représentation graduée, par apprentissage non supervisé hybride entre clustering et analyse factorielle oblique, sous la métrique de Hellinger, afin d'obtenir une image nuancée des thèmes traités : les individus statistiques sont les 212 segments de folio du codex, et leurs attributs sont les 1687 bigrammes de signes extraits. Pour comparaison, nous avons introduit dans cette approche endogène un élément exogène, la décomposition en éléments des signes composites, pour préciser plus finement les contenus. La rétro-visualisation dans le texte original des résultats et expressions dégagées éclaire la signification de certains glyphes peu compris, en les situant dans des contextes clairement interprétables.

1 Introduction et problématique

L'écriture logossyllabique des anciens Mayas, en usage pendant plus de 13 siècles, nous est parvenue au travers de riches inscriptions sur des monuments, des céramiques et trois almanachs divinatoires, qui constituent néanmoins un volume faible de textes disponibles : trois manuscrits et quelques milliers d'inscriptions courtes découvertes.

L'objectif du travail présenté ici est de dégager les principaux contextes sémantiques d'usage des glyphes, dans l'esprit de la sémantique des prototypes (Rosh, 1975), de façon à mettre en contexte commun des glyphes élucidés et ceux qui le sont moins ou pas du tout. A terme, il pourrait déboucher sur la mise à disposition de la communauté scientifique mayaniste de