

Le projet Rosette,

ou l'histoire d'un rêve devenu réalité

par Vincent EUVERTE, texte et photos

De la conception du projet

De plus en plus frustré au retour de chaque voyage en Egypte de ne pouvoir comprendre tous les hiéroglyphes omniprésents sur les murs, stèles, papyri et autres obélisques, Vincent (alias Gardiner sur la Toile) se décide à les étudier en rentrant d'un troisième séjour, au printemps 2004.

Mais apprendre seul, en autodidacte, une écriture d'une telle complexité allait vite s'avérer ardu. Armé de bons ouvrages, et disposant de quelque compétence informatique, il développe rapidement des petits outils d'assistance sous Excel et Access, mais cela ne suffit pas. La technique, c'est bien, mais ça n'est jamais qu'un outil ! Et si d'autres étaient dans le même cas ?

En se rassemblant, ils progresseraient sans doute bien plus rapidement !

Convaincu de ne pas être seul à se débattre dans cette quête, Vincent commence donc à explorer la Toile, et découvre l'existence de multiples sites et fora sur l'égyptologie. Postant de-ci, de-là des messages expliquant son idée naissante d'une assistance informatique à la lecture et à la traduction des hiéroglyphiques, gratuite et libre de droits sur Internet. La réponse ne se fait pas attendre : plutôt que de s'appuyer sur des sites généralistes, pourquoi ne pas créer un site et un forum dédiés à ce projet. Sitôt dit, sitôt fait, Stéphanie et Horus les mettent en place en quelques heures ... Rosette est né le 18 août 2004 ... Vincent n'en est pourtant qu'au chapitre 8 du Grandet & Mathieu, autant dire encore un piètre débutant !

Pourquoi et comment ?

Le nom du projet est venu naturellement : la Pierre de Rosette fut la clef essentielle qui permit à Jean-François Champollion de déchiffrer ces hiéroglyphes dont la compréhension avait été perdue depuis quatorze siècles.

Mais que peut apporter la technologie dans la lecture d'une langue ? Les

traducteurs automatiques de langues vivantes ne sont guère convaincants, alors comment pourrait-on prétendre à un quelconque résultat sur une écriture vieille de 4,000 ans ?

La réponse est pourtant claire

Reste à trouver les moyens de mettre tout cela en musique. Un premier essai sous MS Access s'avère vite impraticable, car très difficile à distribuer. Nous nous orientons donc rapidement vers une version Internet, accessible à tous, en excluant a priori tout téléchargement chez le visiteur. Avec l'appui de certains membres férus d'informatique, le projet commence à prendre forme, même si certains thèmes avancent moins vite que d'autres. Les langages HTML, Java Script, PHP, MySQL, ... deviennent leurs armes quotidiennes. En juillet 2005, moins d'un an après le lancement, la première application est en ligne.

Quelques aspects techniques

Rosette, par le large périmètre de ses objectifs, se doit de rassembler de nombreuses compétences, dans des domaines fort variés. Personne ne peut prétendre maîtriser autant de domaines à la fois ... l'"honnête homme" du XVIIème siècle n'est plus ! Aussi devons-nous faire confiance à chaque membre de l'équipe dans sa spécialité. La reconnaissance des hiéroglyphes reste un sujet en souffrance par manque de ressource. Pourtant, nous

E9 G43 D36 D36 X1 F44 Y1 Z2	<i>iw^s.t</i>	nc.
E9 X1 G43 F44	<i>iw^s.t</i>	nc.
E9 X1 F44 Y1	<i>iw^s.t</i>	nc.
E9 D36 X1 F44	<i>iw^s.t</i>	nc.
E9 G43 D36 D40	<i>iw^s</i>	vb.3.lit
E9 D36 M17 M17 X1 F44 A1 Z2	<i>iw^sy.t</i>	nc.



et simple, et l'a toujours été : Rosette ne prétend certes pas offrir une traduction, mais simplement assister le lecteur partout où cela est possible : de la reconnaissance des caractères par OCR à l'analyse syntaxique, en passant par une représentation graphique de qualité, un dictionnaire de dizaines de milliers de mots, et un corpus de textes suffisamment riche pour permettre des analyses lexicographiques à travers les époques, les supports, les écoles,...

sommes convaincus qu'il est possible de partir d'une photographie numérique ou d'un document scanné et obtenir une identification d'un hiéroglyphe (ou peut-être quelques variantes). La technologie de l'OCR (Optical Character Recognition = reconnaissance optique des caractères) a maintenant fait ses preuves, mais l'écriture hiéroglyphique pose quelques problèmes bien spécifiques : contrairement à la plupart des écritures qui sont linéaires, elle dispose les caractères en deux dimensions ("cadre"), avec des tailles variables, et quatre directions (droite-gauche, gauche-droite, horizontal, vertical). Il nous faut donc reprendre les concepts de l'OCR à la base : localisation spatiale de chaque caractère, normalisation de sa taille, comparaison avec une collection de référence nettement plus nombreuse que dans la plupart des écritures, ...

La codification d'un texte hiéroglyphique en vue du son

exploitation par ordinateur a été fort bien résolue par le Manuel du Codage dès 1984 (Fondation Hugot du Collège de France).

Cette référence peut être considérée comme pratiquement universelle, même si certains logiciels en ont compensé les faiblesses par des légères déviations. Rosette participe actuellement à un projet international de révision de ce Manuel (publication prévue en 2007), et bien évidemment s'astreint à en respecter les règles.

L'affichage graphique des hiéroglyphes fut l'un de nos plus gros soucis durant ces deux premières années : d'abord le passage d'une image digitalisée de chaque hiéroglyphe (typiquement en format GIF) à une police vectorielle (fichier TTF), puis la non disponibilité d'une telle police libre de droits de publication, la complexité de gérer plusieurs fichiers de police pour couvrir les centaines de glyphes nécessaires, ... Finalement, nous avons opté pour deux axes majeurs : développer nos propres fontes vectorielles et les construire au format de l'avenir : Unicode. Nous participons d'ailleurs au projet d'incorporation des hiéroglyphes dans la norme internationale Unicode. Le stockage des informations (catalogue des hiéroglyphes, dictionnaire, corpus de textes, ...) étant un élément essentiel

du projet, il nous fallait prévoir une structure capable d'accueillir de grands. Nous nous sommes rapidement orientés vers les techniques de bases de données, permettant de gérer à la fois des quantités illimitées d'enregistrements, dans autant de langues que nécessaire, de lier les informations entre elles, et d'assurer la portabilité d'un environnement informatique à un autre. Les langages PHP et SQL disponibles sur Internet nous ont apporté une réponse pleinement satisfaisante, garantissant l'évolutivité de l'application.

Rosette se voulant international, il nous fallait gérer un nombre potentiellement croissant de langues. Notre équipe étant au départ francophone, nous avons opté dès le départ pour une version à la fois française et anglaise. Une version arabe est en cours de construction, et nous cherchons des traducteurs pour l'italien, l'allemand et l'espagnol.

<p>Gardiner : A3 EGPZ : 57346 GlyphBasic : 1-36 Transliteration : <i>hms</i> Copyright :</p>	 <p>© Rosette-Mury</p>	<p>Sens s'asseoir, siéger</p> <p>Description Homme assis, genou à terre (var.A7,A17,A17a,B4)</p> <p>Comment Déterminatif dans <i>hmsi</i> 'être assis'</p> <p>Source GM-03 -</p>
 <p>© Rosette-Ptah</p>	 <p>© InScribe</p>	<p>Meaning Sit</p> <p>Description Man sitting on heel</p> <p>Comment Det. in <i>hmsi</i> (hier.A17,B4)</p> <p>Source Gard-01 - 442</p>

Le catalogue des hiéroglyphes est un élément essentiel de l'application. Il permet de faire la relation entre le codage MdC d'un texte, le code Gardiner et la représentation graphique (valeur Unicode du hiéroglyphe). Nous l'avons enrichi d'informations lexicographiques, et sommes en train de construire une source documentaire paléographique qui rassemblera des images de chaque caractère à travers les époques, les graphies, les supports, ... Notre dictionnaire, pour l'instant modeste (2500 mots), va s'enrichir progressivement. Un contact a été pris avec le Griffith Institute pour obtenir l'autorisation de publier la liste des mots clef du dictionnaire de Raymond O. Faulkner. A terme, notre objectif est de fournir une analyse lexicale basée sur une collection de plus de 40,000 mots, documentés et référencés à l'oeuvre d'origine. L'analyseur syntaxique sera certainement le fleuron de l'application. Actuellement en cours de développement, il utilise la technique d'intelligence artificielle nommée

"réseau de neurones".
L'idée de base est de distinguer chaque règle grammaticale et de l'appliquer au contexte de la séquence hiéroglyphique en tenant compte des résultats de l'analyse lexicale.

Le résultat devrait être une série de scénarios possibles, parmi lesquels le lecteur devra faire son choix ... mais cela n'est-il pas la tâche la plus complexe, la plus noble et la plus intéressante pour l'épigraphiste ?

Comme ce qui précède en a jeté les fondations, nous commençons à accumuler des textes translittérés et traduits, dans plusieurs langues, en versions hiéroglyphique, MdC, translittérée, et traduite en plusieurs langues. Et dans la mesure où elles sont disponibles, nous offrons l'image du document d'origine en haute résolution. Regardant l'avenir, le site Rosette s'efforce d'être évolutif et conforme aux normes les plus rigoureuses. Aucun texte n'est publié sans l'accord formel de l'auteur ou du traducteur. Le projet étant totalement bénévole, tout y est gratuit et sous licence GNU-GPL. Les pages sont certifiées XHTML 1.0. Il n'y a aucune action sur l'ordinateur du visiteur (aucun cookie),

tout étant géré par des sessions sur le serveur de Rosette ... Bref, Rosette s'efforce d'être irréprochable, gage de sa crédibilité et de son devenir.

Quelques barrières !

The screenshot shows the Rosette website interface. At the top, there is a navigation menu with tabs: General, Rosette, Quelques plus, Documentation, and Projet. Below the menu, there is a search bar and a quote: "Ils ne savent pas que c'était impossible, alors ils l'ont fait (Mark Twain)". The main content area displays a list of hieroglyphs with their transliterations and a small image of a hieroglyphic fragment. Below the list, there is a section titled "Bienvenue sur l'espace du projet Rosette." and "Quelques News" with a link to "Rosette dans Toutankamon Magazine!".

Les détracteurs du projet ne manquèrent pas dès les premiers jours, et certains y restent encore farouchement opposés ! "Ca ne marchera jamais !" ; "Ca existe déjà !" ; "Vous perdez votre temps !" ...

De quoi en décourager plus d'un, mais les chacals aboient et la caravane Rosette poursuit son bonhomme de chemin. La traversée du désert fut courte, et la suite vous montrera que nous avons eu raison de persévérer.

Une autre difficulté s'est présentée pour l'affichage informatique des hiéroglyphes. Bien que communément utilisées sur Internet, les polices vectorielles disponibles sont pratiquement toutes sous copyright, et nous avons essuyé le refus catégorique de les utiliser dans le cadre de Rosette, à l'exception de celle de Saqqara Technology (logiciel Inscribe) que nous remercions au passage. Ceci nous a conduit à redessiner nous-mêmes non pas une mais deux fontes hiéroglyphiques de haute qualité graphique. Enfin, et c'est bien là notre plus grand souci actuel, il est très difficile d'obtenir le droit de diffusion des traductions existantes, la plupart des auteurs étant farouchement attachés à leur œuvre. Ceci ralentit considérablement la croissance du corpus, et retarde d'autant son exploitation vraiment efficace qui nécessite un volume conséquent de documents. Heureusement, quelques uns ont compris

l'intérêt du projet et contribuent activement à l'enrichissement de notre base de textes.

Une équipe et un réseau



Le monde virtuel d'Internet est surprenant, surtout quand on n'en cible que le bénévolat. Rosette n'ayant aucun objectif mercantile, seul le volontariat à titre gratuit y est admis. La contrepartie en est que chacun peut entrer et sortir à son gré, contribuer selon ses

compétences, ses goûts et ses disponibilités, sans aucun engagement de quiconque. Ceci rend la gestion du projet difficile, mais donne également un état d'esprit unique à l'équipe. La confiance et le respect mutuel sont nos règles d'or. L'équipe Rosette est constituée en moyenne d'une douzaine de membres vraiment actifs, avec un taux

de remplacement d'environ 30% par an. Mais même ceux qui sont partis restent "propriétaires" de Rosette. Le site est le résultat d'une combinaison de compétences que personne individuellement ne pourrait maîtriser, tant les techniques auxquelles il fait appel sont multiples et variées. En support de notre équipe, plus de 230 membres qui nous apportent leur soutien, une idée ou une assistance ponctuelle. A titre d'exemple, le dessin du nouveau site nous a été "donné" par un infographiste professionnel, qui n'avait aucun attrait pour l'égyptologie, mais simplement parce qu'il trouvait que le projet avait "une certaine noblesse et méritait d'être encouragé" ...

Enfin, et ce n'est pas le moindre de nos supports, la communauté des égyptologues amateurs de la Toile nous a généralement bien accepté, et souvent encouragé. Pour preuve, le nombre de sites et fora qui nous référencent, nous envoyant ainsi davantage de visiteurs chaque jour, parmi lesquels trouverons-nous peut-être les futures recrues de l'équipe.

Une équipe communicante

L'un des secrets de Rosette tient dans sa faculté à communiquer et à réagir très vite. Pour cela, nous utilisons tous les moyens modernes à notre disposition, et il est vrai qu'Internet offre désormais des outils particulièrement bien adaptés à nos besoins : le forum pour capitaliser les informations et les références, les "chats" sur MSN Messenger et AIM pour échanger en direct et à plusieurs,



la messagerie personnelle pour les contacts plus individuels, les listes de diffusion pour la distribution de masse, ...

En plus de cette communication virtuelle, nous organisons des rencontres de visu en région parisienne, pour ceux qui s'y trouvent ou y sont de passage.

Apprendre tout en construisant

C'est là une autre des singularités de Rosette : chacun, même un débutant, peut progresser, et peut-être même plus vite qu'avec un livre ou un cours, tout en participant à l'élaboration du projet, voire en lui amenant une

perspective à laquelle l'équipe n'avait pas encore pensé.

Quelques exemples marquants :

Qui peut, mieux qu'un débutant, définir le jeu de questions/réponses à l'attention d'un visiteur qui n'y connaît rien ? Bien sûr, il faudra l'aide de quelques membres plus savants pour valider les réponses, mais leurs futures pages "les

découvert avec Rosette les mystères du langage PHP. En deux mois, il a su les maîtriser si bien qu'il créait l'éditeur actuel du site, dont il peut (et avec eux) être particulièrement fier.

Et les professionnels dans tout cela ?

Rosette est pour l'instant essentiellement constitué d'amateurs. Pourtant, nos objectifs sont plus ambitieux, et peuvent apporter des réponses aux égyptologues de profession. Invité à la conférence internationale "Informatique et Egyptologie" d'Oxford en août dernier, Vincent a pu présenter le projet et a reçu un accueil plutôt

positif, avec des axes clairs de progrès pour mieux être en adéquation avec les attentes de la profession. C'est là un signe de reconnaissance qui nous rend particulièrement fiers de ce que nous avons accompli en deux ans.



Hiéros pour les Nuls" devraient apporter satisfaction à de nombreux néophytes de tous âges, tout en utilisant la puissance de l'application et la rigueur de leur approche.

Bernard, notre graphiste chef, a dû se plonger dans les techniques de vectorisation pour créer les deux superbes fontes Rosette-Ptah et Rosette-Mury, techniques qu'il maîtrise désormais remarquablement.

Benjamin et l'équipe TIPE, soutenus par leur professeur de mathématiques, ont découvert les arcanes de la logique floue pour amorcer le moteur de l'analyse syntaxique, basée sur un concept d'intelligence artificielle nommé "réseau de neurones".

Paul, lycéen de quinze ans, a

Mais au-delà de cet encouragement, nous y voyons une opportunité pour rapprocher les deux mondes amateur et professionnel, trop souvent ignorants l'un de l'autre. A l'instar de l'astronomie où de nombreuses découvertes sont faites par les "dilettantes", et acceptées par les "scientifiques", nous sommes convaincus que chaque visiteur Internet peut contribuer à l'enrichissement d'une base de connaissance commune, gratuite et ouverte à tous, pour le bien de toute la communauté égyptologique.

Quel futur pour Rosette ?

A la vitesse à laquelle nous progressons, nous devrions rapidement disposer d'un dictionnaire digne de ce nom et d'un corpus de textes suffisamment étoffé pour permettre des premières analyses lexicographiques. Comptant sur un effet boule de neige, nous espérons bien que les traductions de textes, les corrections et l'enrichissement du dictionnaire, et l'élaboration de l'analyse syntaxique vont faire de Rosette un outil vraiment performant et utile à tous.

A plus long terme, nous aimerions bien qu'un organisme officiel reprenne leur concept dans un environnement plus professionnel ... mais cela, nous aurons l'occasion d'en reparler quand nous aurons un peu grandi !