

## RÉVISER LE CODAGE DE L'ÉGYPTIEN ANCIEN

Vers un répertoire partagé des signes hiéroglyphiques

[Stéphane Polis](#), [Serge Rosmorduc](#)

Lavoisier | « Document numérique »

2013/3 Vol. 16 | pages 45 à 67

ISSN 1279-5127

ISBN 9782746246430

DOI 10.3166/DN.16.3.45-67

Article disponible en ligne à l'adresse :

-----  
<https://www.cairn.info/revue-document-numerique-2013-3-page-45.htm>  
-----

Distribution électronique Cairn.info pour Lavoisier.

© Lavoisier. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

---

# Réviser le codage de l'égyptien ancien

## Vers un répertoire partagé des signes hiéroglyphiques

Stéphane Polis<sup>1</sup>, Serge Rosmorduc<sup>2</sup>

1. Fonds National de la Recherche Scientifique – Université de Liège  
Département des Sciences l'Antiquité  
Place du XX Août, 7  
B-4000 Liège, Belgique  
s.polis@ulg.ac.be

2. Cédric/ILJ/Conservatoire National des Arts et Métiers  
Rue Conté, 2  
75003 Paris, France  
serge.rosmorduc@cnam.fr

---

*RÉSUMÉ.* Nous proposons de réviser le codage de l'égyptien ancien qui repose sur un standard nommé « manuel de codage » (1988) ne répondant pas aux besoins actuels dans la création de corpus hiéroglyphiques. Une analyse des 60 000 graphies du corpus Ramsès nous permet de faire deux propositions concrètes concernant, d'une part, les principes présidant à l'encodage des graphies hiéroglyphiques dans les corpus annotés et, d'autre part, la nécessaire refonte du répertoire des signes hiéroglyphiques.

*ABSTRACT.* Based on an analysis of the 60 000 spellings found in the Ramses corpus, we show that the encoding scheme that is used for hieroglyphic texts, known as the Manuel de Codage (1988), is problematic for the development of text corpora in general. We consequently (1) argue in favour of basic principles that should be followed when encoding hieroglyphic texts in the framework of corpus projects, and (2) suggest guidelines for a new encoding scheme, especially with respect to the structure of the sign-list.

*MOTS-CLÉS :* hiéroglyphes, encodage, Unicode, liste de signe, égyptien ancien.

*KEYWORDS:* hieroglyphs, encoding, Unicode, sign-list, Ancient Egyptian.

---

DOI:10.3166/DN.16.3.45-67 © 2013 Lavoisier

Dans cette contribution, nous proposons de réviser le codage de l'égyptien ancien qui repose sur un standard nommé « manuel de codage » (1988). Ce dernier ne répond pas aux exigences contemporaines en matière de représentation informatique des écritures, en particulier parce que (1) la liste des signes qui y sont présents n'est nullement documentée et parce que (2) les opérateurs de positionnement relatif des signes manquent de cohérence dans leur syntaxe et leur sémantique. Nous traitons ici uniquement des problèmes posés par le premier point, c'est-à-dire le répertoire des signes hiéroglyphiques. En effet, en l'état, il entrave directement la bonne constitution de corpus annotés de la langue des pharaons (voir déjà Nederhof, 2013, avec références).

Après une introduction générale concernant les systèmes d'écriture employés en Égypte pharaonique dans laquelle on montre que les degrés variables de normalisation de l'écrit dans les publications égyptologiques ont assez directement influencé la constitution des listes de signes hiéroglyphiques existantes, le texte est organisé en trois sections. Dans la première, nous présentons les principes présidant actuellement au codage de l'écriture hiéroglyphique et décrivons plus avant les problèmes posés par les répertoires de signes existants, notamment dans le standard Unicode. Une deuxième section est consacrée à l'analyse des sources de la variation graphémique dans le corpus annoté *Ramsès*, qui contient environ 430 000 mots et quelque 60 000 graphies différentes. À partir de ces observations, une troisième partie est consacrée à une proposition concrète en vue de l'encodage des graphies hiéroglyphiques dans les corpus annotés. En guise de conclusion, nous élargissons le propos en suggérant un nouveau principe d'organisation du répertoire des signes hiéroglyphiques dans son ensemble. Cette refonte du système de codage de l'écriture égyptienne devrait permettre de construire progressivement un répertoire raisonné de l'ensemble des signes hiéroglyphiques qui soit référencé, décrit et explicitement hiérarchisé.

### Note terminologique

Dans cet article, nous utilisons la terminologie suivante :

– « graphie » désigne la forme graphique d'un mot, telle qu'elle est rencontrée dans un document ou un ouvrage, sans préjuger de son codage ultérieur. C'est une orthographe du mot dans un certain contexte ;

– « signe » est employé de manière générique, à la fois pour l'artefact archéologique, sa représentation, ou le symbole dans une police hiéroglyphique. Le terme nous évite en particulier d'avoir à décider prématurément du statut de ce que nous codons ;

– « caractère » et « glyphe » ont leur sens usuel lorsque l'on parle de codage de textes.

## 1. Introduction

Avant l'apparition de l'écriture démotique (dans le courant du VII<sup>e</sup> siècle avant notre ère), deux systèmes d'écriture ont été employés en Égypte ancienne : l'écriture hiéroglyphique et l'écriture hiératique (Winand, 2013). Attestés à partir de la fin du IV<sup>e</sup> millénaire, ces deux systèmes sont intimement liés. En effet, le second est une forme cursive (généralement tracée à l'encre) du premier, qui est essentiellement employé dans des contextes monumentaux (avec des signes qui sont gravés ou peints sur les parois) (Vernus, 1990).

Si la tachygraphie hiératique participait de la formation de base des scribes, l'apprentissage de l'écriture hiéroglyphique semble avoir été réservé à une autre phase de l'enseignement. Par conséquent, cette dernière était maîtrisée par un nombre plus restreint de lettrés. Quoi qu'il en soit, ces deux écritures sont faites de signes connus pour entretenir un lien plus ou moins évident (c'est-à-dire moyennant les principes de figuration propre au dessin égyptien) avec un référent extralinguistique. Ainsi, la lettre *m* peut-elle s'écrire avec le signe de la chouette, dont on reconnaîtra la stylisation aussi bien en hiéroglyphe qu'en hiératique :



Figure 1. La lettre *m* en hiéroglyphe et en hiératique

Étant donné les relations iconiques perceptibles entre ces deux systèmes d'écriture, les égyptologues ont pris l'habitude de normaliser dans leurs éditions tous les signes – qu'ils soient originellement hiératiques ou hiéroglyphiques – en recourant à l'écriture hiéroglyphique, plus directement lisible. On mesure aisément l'ajout et la perte d'information qui résultent d'un tel processus de standardisation : les deux *m* de la figure 1 auront ainsi le même code, enrichissant et appauvrissant, tant en quantité qu'en qualité, les informations véhiculées par chacun des systèmes graphiques.

Ce point est crucial pour comprendre l'évolution du répertoire des signes hiéroglyphiques sur lequel reposent les fontes existantes, car l'attitude face à la standardisation des signes a varié en fonction du système d'écriture. En effet, si la normalisation généralisée des signes de l'écriture hiératique en signes hiéroglyphiques semble n'avoir jamais ému personne, sanctionnant ainsi une approche typographique par caractère (ou grands types de l'écriture hiératique), les égyptologues ont en revanche été particulièrement attentifs à développer des fontes qui soient aussi fidèles que possible à l'original pour l'édition imprimée de documents hiéroglyphiques. C'est donc en tenant compte d'un luxe de détails que les listes de signes hiéroglyphiques se sont enrichies à mesure que les fontes grandissaient (approche

bien sûr légitime dans une approche paléographique, cf. Meeks, 2007), sans que soient en parallèle documentés les emplois de chacun des signes ou que l'on sache s'il s'agit d'une variante particulière à un lieu, voire à une époque, ou bien d'un nouveau caractère non encore répertorié.

En vue de l'*impression* d'éditions de textes, cette approche n'est pas problématique : il est légitime qu'un éditeur souhaite rendre avec autant de précision que possible l'original qu'il a à publier. Comme nous allons le voir, il n'en va toutefois pas de même lorsque ce sont ces mêmes fontes hétérogènes qui sont utilisées pour servir de base au *codage*, c'est-à-dire la représentation informatique, de l'écriture hiéroglyphique.

## 2. Représentation informatique des hiéroglyphes

Les premières tentatives de codage informatique de l'égyptien datent de la fin des années 1960. Pour les textes hiéroglyphiques, le premier système fonctionnel fut probablement *Glyph*, développé sur gros systèmes par l'astrophysicien Jan Burman (Burman, 1976). Durant les années 1980, avec l'apparition des ordinateurs personnels, les premiers logiciels d'édition hiéroglyphique furent diffusés, la plupart visant uniquement à permettre l'insertion de hiéroglyphes dans des publications papier. L'idée de réaliser, dans le futur, des corpus électroniques était cependant déjà très présente. En 1984, le groupe de travail « Informatique et Égyptologie » de l'Association Internationale des Égyptologues se réunit pour la première fois, et décida la mise au point d'un format de description des textes hiéroglyphiques. Ce format fut fondé sur celui utilisé par le logiciel *Glyph*, mais en étendant notablement la collection de signes d'origine. En 1986, paraissait la première édition de « l'inventaire des signes hiéroglyphiques en vue de leur saisie informatique », appelé communément « manuel de codage » ou « MdC » (Burman *et al.*, 1988). Le système décrit dans le « manuel » est aujourd'hui utilisé par la plupart des traitements de textes hiéroglyphiques, avec diverses extensions et modifications (Gozzoli, 2013).

### 2.1. Les principes du manuel de codage

Pour décrire correctement un texte hiéroglyphique, il ne suffit pas de décrire la séquence des signes qui le composent. En effet, les hiéroglyphes occupent l'espace d'une manière complexe. Un système de codage des textes hiéroglyphiques doit donc comporter à la fois un codage des signes et des opérateurs de positionnement.

Pour prendre un exemple, le mot « ciel », *p.t* en égyptien, comporte les trois signes □ (*p*) ◊ (*t*) et ≡ (signe représentant le ciel). Ils peuvent être disposés ainsi : □◊, □◊ ou ◊□. Le choix de la disposition des signes dépend d'une multitude de facteurs : habitudes graphiques, support, espace disponible, coupure des mots, etc.

Le *manual de codage* n'utilise que des caractères ASCII. Les hiéroglyphes sont codés, soit par leur valeur phonétique pour une partie d'entre eux (ainsi, le  peut être codé « p » et  peut être codé « pt »), soit par des codes inspirés de ceux qui étaient utilisés pour repérer les signes dans les polices en plomb, et plus précisément par les codes de la police développée pour l'édition de la grammaire de sir Alan Gardiner (1957). Ces codes sont composés d'une lettre, qui décrit le type d'objet représenté par le hiéroglyphe (A pour « homme », B pour « femme », ... O pour « bâtiment », Z pour « traits », et Aa pour « non identifié »), et d'un numéro. Ainsi «A1» code le signe , «N1» est un autre code pour .

Les opérateurs de positionnement permettent de former des « cadrats », carrés idéaux selon lesquels les signes sont disposés. Ainsi, dans un texte en ligne, “-” permet de séparer deux cadrats, “:” permet de superposer verticalement deux signes, et “\*” permet de les combiner horizontalement. Le système permet aussi d'user de parenthèses pour réaliser des groupements plus complexes. Pour reprendre notre exemple précédent, le mot ciel peut donc se coder : « p\*t:pt » et il sera alors rendu .

À l'usage, les opérateurs de positionnement proposés en 1984 ne permettent qu'une approximation assez grossière des textes. Plusieurs extensions non standard ont été mises en œuvre dans les éditeurs de textes hiéroglyphiques. La proposition la plus aboutie étant celle de Nederhof (2002 ; 2013), la seule qui fournisse actuellement une syntaxe précise et une sémantique claire.

Néanmoins, comme on l'a signalé plus haut, le problème le plus important réside dans le manque complet de documentation et de structure de la liste des signes hiéroglyphiques.

## 2.2. Le répertoire des signes

Les premières polices hiéroglyphiques datent du XIX<sup>e</sup> siècle (Janssen, 1972), avec notamment la police Theinhardt, développée sous l'impulsion de l'égyptologue allemand Richard Lepsius. Au XX<sup>e</sup> siècle, à côté de très nombreuses éditions autographiées, on a utilisé principalement deux polices. La première, qui comporte un peu plus de 800 signes, a été créée pour l'édition de la grammaire de A. H. Gardiner (1957), et dessinée par les époux de Garis Davies (Gardiner 1963, p. xv). La police Gardiner a connu par la suite quelques extensions mineures. Elle couvre un « noyau dur » des hiéroglyphes, mais s'avère insuffisante pour l'édition de la plupart des textes monumentaux. L'Institut français d'archéologie orientale, de son côté, a très tôt développé une police riche de plus de 7 000 signes, utilisée notamment pour l'édition des temples de l'époque gréco-romaine. Cette richesse est cependant trompeuse (Collombert, 2007 ; Meeks, 2007), car la palette de signes a été étendue en fonction des besoins ponctuels d'édition et non d'une vision d'ensemble du système hiéroglyphique, incluant maintes variantes stylistiques.

Quant au répertoire des signes du *manuel de codage*, qui nous intéresse ici au premier chef, il a été développé à partir des polices existantes (essentiellement les fontes Gardiner, IFAO et Theinhardt) et des fiches ayant servi à la rédaction du dictionnaire de Berlin (Erman et Grapow, 1926-1963).

On y trouve à l'origine une distinction similaire à celle que fait Unicode entre caractère et glyphe : *On (...) distingue strictement graphèmes et variantes graphiques, les graphèmes étant [des] hiéroglyphiques qui se distinguent les uns des autres par leur lecture et leur sens* (Buurman et al. 1988, p. 51). Les *variantes graphiques* sont signalées par un code qui se termine par une lettre. A1A est donc une variante de A1. Dans les faits, cette distinction est toute théorique.  (M1B) est bien marqué comme une variante graphique de  (M1), mais  (G208), qui n'est qu'une variante de  (G21), a un code de « graphème » selon la terminologie du *manuel de codage*. La hiérarchisation à deux niveaux entre graphèmes et variantes (correspondant *mutatis mutandis* au couple Unicode caractère-glyphe), qui demandait une recherche approfondie sur le fonctionnement des très nombreux signes de l'écriture hiéroglyphique, n'a donc jamais été menée à bien.

On pourrait multiplier de tels exemples à l'envi, mais à vrai dire, ces incohérences dans la hiérarchisation ne sont probablement pas le principal problème du répertoire de signes du *manuel de codage* en l'état. L'absence presque complète de documentation porte bien plus à conséquence. Rien n'indique d'où proviennent les signes, ni la raison pour laquelle ils sont inclus dans la liste. Il est certain que de telles informations eussent été très longues à réunir, en particulier pour les signes rares repris directement aux polices anciennes. Néanmoins, cela pose de nombreux problèmes. Les créateurs de nouvelles polices hiéroglyphiques, tout d'abord, ne disposant en tout et pour tout que d'une image du signe, ne savent *a priori* pas quels en sont les traits distinctifs ; comme, dans une police, il importe d'harmoniser la représentation graphique des signes, l'absence de description complique le travail du typographe. Le problème se pose aussi à l'utilisateur/encodeur. Pour un hiéroglyphe original sur un monument ancien, il arrive assez fréquemment que les polices proposent plusieurs signes visuellement proches, sans qu'aucun ne coïncide totalement avec celui qu'il a sous les yeux. Sans plus de documentation disponible, l'encodeur finit souvent par faire un choix arbitraire qui peut être lourd de conséquences dans le cadre de la constitution de corpus annotés.

### 2.3. Évolution du répertoire de signes du manuel de codage

Au cours des années 1990, le destin du *manuel de codage* fut lié à celui des deux logiciels *MacScribe* et *Winglyph*, diffusés par le *Center for Computer-aided Egyptological Research* (CCER), localisé à l'université d'Utrecht. Les polices furent étendues, en collaboration avec l'Institut français d'archéologie orientale, pour l'édition d'un certain nombre de monuments, en particulier les chapelles osiriennes du temple de Dendera (Cauville, 1997). Le principal créateur des polices, l'égyptologue J. Hallof, et le créateur du logiciel *Winglyph*, H. van den Berg, développaient

en parallèle une base de données, le système *Toth* (Hallof et van den Berg, 1992). La saisie des textes de Dendera leur permit ainsi de produire un index exhaustif des mots égyptiens attestés dans les chapelles osiriennes.

Ce mode de fonctionnement, le seul qui permettait à l'époque d'assurer le financement du développement du système, déséquilibre le répertoire. Les signes tardifs de certains sites y sont représentés avec un grand luxe de détails, alors que certains signes importants d'époques antérieures manquent.

#### 2.4. Unicode

Le standard Unicode comporte un codage des hiéroglyphes depuis sa version 5.2 (Everson, 2006). Par rapport aux propositions initiales, cette version du standard ne comporte qu'une sélection restreinte de signes, suffisante pour écrire, par exemple, une grammaire, mais peu adaptée à la constitution de bases de données linguistiques. Le modèle même d'Unicode, supposant un inventaire fermé des caractères, est d'ailleurs en porte-à-faux avec le côté ouvert de l'écriture hiéroglyphique (Bunz, 2000 ; Rosmorduc, 2002).

Pour ces raisons, le Projet Ramsès s'est vu obligé de recourir, dans un premier temps au moins, au répertoire de signes du *manuel de codage*<sup>1</sup> (voir ci-dessous).

Nous utilisons l'encodage hiéroglyphique d'Unicode de manière épisodique : un prototype de dictionnaire hiéroglyphique sous Android, fondé sur notre lexique, a été développé ; une exportation « dégradée » des graphies permettra éventuellement une conversion au format de la TEI (supposant un codage Unicode), exploitable sans qu'il soit besoin de disposer d'un moteur de rendu spécialisé ; de même, Unicode permet d'exporter de manière relativement lisible des résultats de recherches vers des tableurs.

### 3. La variation des graphies dans le projet Ramsès

Le projet Ramsès vise à la construction d'un corpus richement annoté des textes écrits en néo-égyptien, et plus largement, de tous les textes produits durant le Nouvel Empire et la Troisième Période Intermédiaire (c. 1350-700 ACN) dont les registres d'expression s'ouvrent à des traits linguistiques caractéristiques du néo-égyptien (voir Polis et Winand, 2013, avec la bibliographie antérieure). Une des originalités du projet consiste en l'encodage systématique des graphies hiéroglyphiques, ce en quoi il se distingue d'autres projets existants, comme le fameux *Thesaurus Linguae Aegyptiae*.

Étant donné la nature du système d'écriture et l'absence complète de *tagger* et de *parser* pour l'égyptien ancien, l'ensemble du corpus est annoté manuellement. Le

---

1. La plupart des bases qui codent des textes hiéroglyphiques font de même (par exemple Gourdon, 2013 ; Hallof et van den Berg, 1992), quand elles ne se contentent pas de translittérations.

travail d’annotation a commencé fin 2006 et, en août 2013, le corpus rassemble environ 3 300 textes et quelque 435 000 occurrences. Plus important pour la présente discussion est le nombre de graphies enregistrées dans notre corpus : environ 60 000 graphies différentes pour 9 000 lemmes.

Nous estimons que ce nombre relativement important de graphies, qui ont l’avantage de provenir de textes appartenant à des genres variés et ayant été inscrits sur des supports divers – aussi bien en caractères hiéroglyphiques que hiératiques – constitue une base solide (même si non idéale, en raison de l’extension diachronique limitée des données) pour revoir le système de codage des hiéroglyphes et, en particulier, repenser l’organisation du répertoire des signes sur laquelle repose ce codage.

### 3.1. Le principe d’encodage des graphies dans Ramsès

Ainsi qu’on l’a évoqué plus haut, les graphies du projet Ramsès reposent sur le *manuel de codage*, avec quelques extensions proposées par le logiciel *JSesh* (en particulier quelques codes supplémentaires proposés par Gardiner (1929), sur la base de ses observations concernant la transcription du hiératique du Nouvel Empire).

Concernant les principes suivis pour l’encodage des graphies, on distinguera trois phases successives.

Durant une période de six années (fin 2006 à 2013), l’encodage des graphies a été *systématique*. On veut dire par là que les personnes en charge de l’encodage avaient pour instruction de reproduire aussi fidèlement que possible *l’édition hiéroglyphique de référence* du texte qu’ils encodaient, sans revenir au document original. Cette approche était justifiée par une volonté de rapidité de codage des textes et permettait aux encodeurs moins expérimentés (notamment des étudiants en thèse) de s’appuyer sur l’autorité des choix opérés dans l’édition de référence. Cette option n’est pas sans poser certains problèmes : les choix de normalisation du hiératique sont variables en fonction des éditeurs, voire chez un même éditeur.

On illustrera ce point avec un exemple. Dans une publication de lettres de la fin de la période ramesside, J.J. Janssen (1991) transcrit le morphème grammatical du conjonctif *mtw* (en hiératique dans l’original), tantôt  , tantôt  . Comme le montre la figure 2, l’éditeur semble choisir de standardiser les originaux hiératiques, très similaires, de deux manières différentes dans son édition autographe. Ce qui pourrait laisser penser que la distinction entre  et  n’est pas significative et que l’un vaut pour l’autre.

Pourtant, sans entrer inutilement dans le détail de la transcription du hiératique, le P. BM EA 10373 reproduit figure 3 possède une autre graphie du morphème grammatical, qui est rendu par l’éditeur de manière parfaitement appropriée,  .

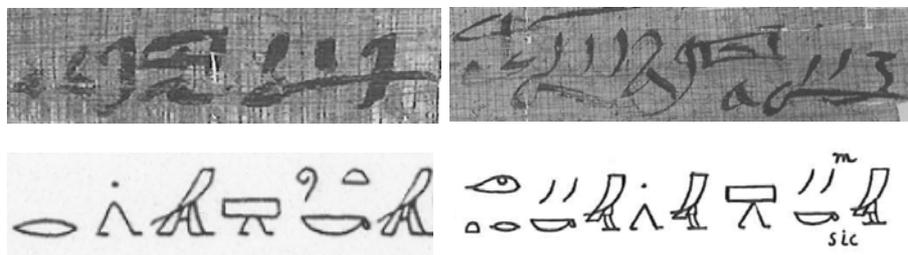


Figure 2. Original et transcription hiéroglyphique du P. BM EA 10412, v° 4 (= Janssen 1991, pl. 41-42) et du P. BM EA 10373, v° 3 (= Janssen 1991, pl. 29-30)

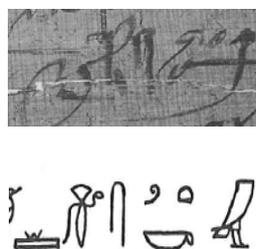


Figure 3. Original et transcription hiéroglyphique du P. BM EA 10373, v° 6 (= Janssen 1991, pl. 29-30)

Dans ce document, l'opposition entre deux transcriptions est donc motivée par une variation dans l'original, alors qu'elle ne l'était pas dans le premier exemple. On mesure aisément que le choix de suivre aussi fidèlement que possible la transcription fournie par une édition de référence a induit des problèmes de cohérence, limités mais réels, dans l'encodage des graphies au sein du corpus Ramsès.

Des problèmes de cet ordre, combinés à ceux que l'on pouvait supputer en raison de l'absence de documentation du répertoire des signes hiéroglyphiques évoquée ci-dessus (section 2.2), nous ont conduit à envisager une seconde étape (2013-2014), à savoir une enquête approfondie, un temps de réflexivité critique, visant à analyser plus en détail les pratiques de notre propre équipe pour le codage des graphies hiéroglyphiques dans Ramsès. Dans la pratique, les observations et propositions formulées dans le cadre de la présente étude préliminaire reposent sur une extraction de tous les codes des signes actuellement utilisés dans le corpus, triés par fréquence d'emploi et liés à la liste des lexèmes dans lesquels ils sont employés. L'analyse du fichier en sortie nous a permis de mettre en évidence divers phénomènes marginaux dans l'encodage des graphies : emplois de signes rares, cas de doublons (par exemple, deux variantes stylistiques pour un même signe), combinaisons *a priori* inattendues de signes, etc.

Au final, cette réflexion sur nos pratiques actuelles devrait conduire à la troisième phase. Celle-ci pourrait commencer dès 2014 et reposer (1) sur une charte précisant les principes devant présider à un encodage non plus *systématique*, mais *raisonné* des graphies hiéroglyphiques dans un corpus annoté – ce point fera l’objet de la section 4 – et (2) sur un répertoire des signes explicitement hiérarchisé et documenté, suivant les propositions faites en section 5.

### 3.2. Esquisse d’une typologie de la variation graphémique

Les incohérences que nous avons pu relever dans l’encodage des graphies hiéroglyphiques ont des origines diverses. Elles peuvent être liées à la variabilité des graphies dans les documents originaux, aux choix posés par les éditeurs ou à des options retenues par les encodeurs. Ces trois sources de variation dans l’encodage des graphies hiéroglyphiques, parfois combinées, vont nous permettre d’enrichir la réflexion visant à définir des principes d’encodage de l’écriture hiéroglyphique.

#### 3.2.1. Le document original

Étant donné que les encodeurs avaient pour recommandation de reproduire d’aussi près que possible les signes rencontrés dans la publication de référence, certaines variantes paléographiques – non significatives au niveau linguistique et aberrantes au niveau du codage – ont fait leur apparition au sein du corpus. On en prendra un exemple particulièrement topique. Une épithète extrêmement fréquente pour les rois et *realia* associés, *nh-wdꜣ-snb* « en vie, force, santé », est normalement écrite  $\text{𓏏} \text{𓏏} \text{𓏏}$ . Le dernier signe est le phonogramme *s*, ici employé comme abréviation de *s(nb)* « santé », qui représente iconiquement un vêtement replié et dont le code est S29 dans le *manuel de codage*. Or on trouve dans Ramsès une graphie  $\text{𓏏} \text{𓏏}$  où la partie gauche du *s* semble avoir été raccourcie. Après analyse, il s’avère que cette graphie a été créée pour rendre compte de la forme particulière du signe S29 dans une inscription de Médinet Habou dont on donne un fac-similé en figure 4.



Figure 4. *Epigraphic Survey* (1932, pl. 75)

Afin de rendre l’original avec un signe *visuellement* proche, l’encodeur a ici choisi de recourir au signe  $\text{𓏏}$  (S39) auquel il a fait subir une rotation en miroir pour

obtenir un rendu  $\text{𓆎𓆏}$ . Ce faisant, le code employé est particulièrement trompeur. En effet, le signe S39 représente le sceptre ou bâton de berger, qui n'a rien à voir avec le signe du linge replié possédant la valeur phonétique *s*.

On conclura que cette vision paléographique et ce codage mimétique n'ont guère lieu d'être dans la constitution d'un corpus annoté. Le contexte ne laisse ici demeurer aucun doute : c'est bien S29 qui était écrit et le signe devrait probablement être codé comme tel dans un corpus annoté. Toutefois, il devrait être parallèlement possible d'enrichir le répertoire des signes de variantes stylistiques ou paléographiques pour autant que ces dernières (1) soient explicitement déclarées comme se rapportant à un type (ici le signe du linge replié fonctionnant comme phonogramme avec valeur *s*) et (2) soient décrites adéquatement. Nous aurons l'occasion de revenir sur ce point plus en détail dans les conclusions.

### 3.2.2. L'édition du texte

Les variations induites par les choix de l'éditeur ont déjà été évoquées en section 2.1. Plus avant, il convient de préciser que les principes d'encodage en corpus devraient probablement dépendre du système d'écriture source. Il faudrait ainsi distinguer le cas de l'écriture hiératique de celui de l'écriture hiéroglyphique.

On prendra un exemple de codage particulièrement extravagant pour illustrer le premier point. Dans un ostracon littéraire de Deir el-Médineh, à la fin du nom du dieu Taténe, le signe représentant ledit dieu a été transcrit par une figure debout portant cornes de bélier et coiffé de plumes (cf. figure 5).

Aucun signe hiéroglyphique ne correspondant à cette forme particulière dans le *manuel de codage*, l'encodeur a opté pour le codage S77\73\*\*A53{{66,713,194}}, qui consiste à placer la couronne S77  $\text{𓆎}$  sur la tête de la figure A53  $\text{𓆏}$ . Or, la paléographie de référence du hiératique (G. Möller 1927, II, 70) précise que la transcription standardisée qui est retenue pour cet original hiératique est le signe C18  $\text{𓆏}$ , avec une représentation assise du dieu en question, qui aurait pu être utilisé en toute légitimité.



Figure 5. O. DeM 1703 & 1704, l. 2 (d'après Gasse 1990, s.n.)

Cet exemple nous permet de conclure que, pour la transcription des textes hiératiques, les encodeurs ne devraient idéalement utiliser qu'un nombre restreint de signes hiéroglyphiques qui soient explicitement définis dans le répertoire comme correspondant potentiellement à une transcription standard du hiératique. Faute d'un tel principe, on imagine aisément les difficultés nombreuses et variées qui se poseront lors des futures recherches sur corpus.

Un principe identique ne vaut pas pour les textes rédigés en hiéroglyphes. Comme on l'a précisé (section 2.4), ce système d'écriture est fondamentalement ouvert aux innovations, qui sont consciemment introduites par des scribes capables de jouer avec toutes les potentialités du système hiéroglyphique.

On en prendra pour exemple une graphie inhabituelle du verbe *sḥ*, qui signifie « dresser, ériger », dans une stèle du roi Ahmôsis de Karnak. Dans la première édition de ce texte fournie par Legrain (1903, 29), la graphie hiéroglyphique proposée est . On ne comprend alors pas pourquoi un verbe signifiant « dresser, ériger » est orthographié avec un classificateur humain relatif à l'adoration ou à la prière (). À y regarder de plus près, l'original est en fait écrit , avec la combinaison du phonogramme *ḥ* et du classificateur relatif à la construction , le scribe ayant pour l'occasion choisit de substituer au dessin du mur celui du phonogramme *ḥ*, de forme verticale similaire, intervenant dans la graphie du verbe *sḥ*. Le jeu graphique est donc bien conscient et significatif : le codage doit pouvoir en rendre compte et prévoir l'enrichissement constant du répertoire avec des signes qui soient décrits aussi précisément que possible au niveau des unités qui les constituent.

Il ne fait par ailleurs pas de doute que c'est une limitation de la fonte disponible au début du XX<sup>e</sup> siècle qui a conduit à la normalisation abusive d'un hiéroglyphe complexe dans l'édition imprimée. On signalera en ce sens que le signe est parfaitement identifié dans l'édition autographe postérieure de K. Sethe (*Urk.* IV, 23,14 : ) et fut introduit dans la liste des signes hiéroglyphiques sous le code A170. Quoi qu'il en soit, on observe à quel point les éditions de référence (typographiées ou autographes) peuvent être source de variation et d'incohérence pour l'encodage des documents hiéroglyphiques et que seule une vérification sur l'original ou des photographies de qualité permettrait de valider ou d'infirmier l'existence de signes rares et complexes permis par le système.

### 3.2.3. Le codage des hiéroglyphes dans Ramsès

L'un des problèmes les plus fréquemment rencontrés, qui est dû, en partie, au manque de documentation des polices existantes, est le choix des signes sur la base de ressemblances graphiques superficielles<sup>2</sup>.

2. Voir déjà les remarques en section 3.2.1.

Pour donner un exemple du type de problèmes potentiels, le répertoire contient plus de cinquante signes représentant un ennemi vaincu, dans des postures diverses :  (A13),  (A13C),  (A422), etc. Malgré cette richesse apparente, il arrive fréquemment qu'aucune des variantes disponibles ne convienne tout à fait. Dès lors, comment choisir le signe le plus pertinent ?

Comme on l'a dit à plusieurs reprises, la consigne donnée aux codeurs fut de rester le plus proche possible de leur source, qui était souvent une édition du texte original, donc déjà une source secondaire, à mi-chemin entre diplomatique et critique (suivant les coutumes éditoriales égyptologiques, sensiblement différentes de celles des humanités classiques<sup>3</sup>). Un inconvénient de ce choix fut qu'il conduisit les codeurs, par scrupule, à tenter de se rapprocher le plus possible du graphisme du signe trouvé dans leur source, au détriment d'une analyse sémiologique.

L'opération de codage est loin d'être simple, et demande souvent une analyse assez poussée du texte. Pour prendre un exemple, dans la liste des signes rares relevés dans la base, nous avons trouvé le signe  (E95) : une vache couronnée d'un disque solaire. Elle déterminait le mot « j3w », qui signifie « bétail ». Dans un premier temps, nous supposâmes un de ces jeux d'écriture dont les Égyptiens étaient friands. Le « bétail du dieu soleil » est effectivement une métaphore usuelle pour l'humanité. Après consultation des sources, il apparut que le bovin portait une coiffure plus complexe :  (KRI II, 37,6). Ce pouvait être, éventuellement, un disque surmonté de deux plumes. Cependant, le contexte contredisait notre première hypothèse : on ne parlait pas là de « troupeau divin », mais de bétail de sacrifice. Le hiéroglyphe E255 du *manuel de codage* pouvaient alors convenir, même si formellement assez éloigné : . Il représente un bœuf gras, orné de guirlandes. Néanmoins, le signe des polices et celui trouvé dans le document diffèrent sensiblement, et, en se fixant uniquement sur le graphisme, un signe comme (E98, une vache portant un disque solaire surmonté de plumes) aurait pu sembler plus proche. On voit la difficulté de choisir le caractère approprié : ici, le choix correct demandait donc de mettre en œuvre, en dehors de la philologie, des connaissances sur l'iconographie du bétail sacrificiel.

Dans d'autres cas, le choix fait, sans causer de vraies erreurs, a conduit à multiplier des variantes graphiques qui n'avaient pas lieu d'être. Pour une bonne partie de nos textes, l'original est écrit en hiéroglyphique. Les codeurs travaillent généralement sur des transcriptions hiéroglyphiques du texte, soit réalisées par l'éditeur de celui-ci, soit trouvées dans de grands recueils comme les *Ramesside Inscriptions* de K. Kitchen, le retour au document original ne se faisant qu'en cas de problème de lecture patent. Or, ces transcriptions sont généralement manuscrites, sans pour autant être des fac-similés. La forme des signes ne dépend alors, sauf exception, que de la « main » de l'éditeur lui-même. Ainsi, pour le signe phonétique

3. On reste proche des graphies originales, en tentant de les respecter au mieux, tout en se permettant, par exemple, de restaurer les lacunes du texte, et d'ajouter des marques ecdotiques.

T22 (𐀓), qui se translittère « sn », on rencontre de nombreuses variantes dans la base – 𐀓 (T23), 𐀓 (T22A) ou 𐀓 (T22B) –, alors même que toutes se résolvent dans le même signe hiéroglyphique dans l'original. C'est d'autant plus gênant que, pour d'autres signes, les variantes rencontrées en hiéroglyphique peuvent être significatives, par exemple 𐀓 (H6) et 𐀓 (H6A). Dans le cas de T22, le choix complique la recherche sur les graphies, et la seule information apportée concerne éventuellement l'identité de l'éditeur du texte. La forme la plus souvent retenue par les éditeurs, T22A (la plus facile à dessiner), est en outre, d'après Meeks (2005, 183, § 497), une forme tardive, et donc *a priori* hors de notre période. Les éventuelles informations fournies par notre base sur les variantes de T22 sont, dans ces conditions, inutilisables.

#### 4. Principes d'encodage des graphies dans les corpus hiéroglyphiques

Les observations qui précèdent nous conduisent à avancer deux propositions concrètes concernant l'encodage des graphies dans les corpus annotés de l'égyptien ancien. Ces propositions visent à permettre une reproductibilité du codage ainsi que l'échange de ressources entre des projets complémentaires en égyptologie. On suggère que les possibilités de codage pour un signe donné dépendront :

- 1) de la fonction du signe au sein du mot dans lequel il apparaît ;
- 2) du système d'écriture du document source (hiéroglyphique ou hiéroglyphique).

##### 4.1. Tenir compte de la fonction des signes hiéroglyphiques

La première proposition que nous faisons repose sur l'hypothèse que, *dans le cadre de corpus annotés*, les très nombreuses variantes potentielles d'un signe n'ont pas la même importance dans tous les cas. Pour le dire autrement, il est des emplois où une variante graphique d'un signe affectera le sens d'un mot (en en modifiant ou précisant le sens en contexte) et d'autres où la variation n'est pas significative au niveau linguistique (même si elle peut évidemment l'être au niveau indexical, paléographique, etc.).

Ce point nous impose de préciser dans un premier temps quelles sont les fonctions possibles pour un signe en contexte<sup>4</sup>. Les grammaires classent généralement les signes (Winand 2013, p. 30 et suivantes) en :

- Idéogrammes (ou logogrammes) : signes-mots.
- Phonogrammes : signes codant des consonnes ou semi-voyelles.
- Déterminatifs (ou classificateurs<sup>5</sup>) : signes écrits à la fin d'un mot, ils précisent le domaine sémantique dont il relève.

4. On donne ici une présentation volontairement simplifiée, mais suffisante dans le cadre de cette contribution. Pour une discussion détaillée, voir Polis et Rosmorduc, à paraître.

Ainsi, le mot « bœuf », *jh* en translittération, peut s'écrire avec l'idéogramme du bovin :  (le trait est un signe diacritique qui marquait, au départ, le caractère idéographique du signe qui le précédait). Le mot serviteur, *b3k*, peut s'écrire  ; les deux premiers signes sont des phonogrammes (l'oiseau se lit *b3*, et la corbeille se lit *k*) ; le dernier signe est un déterminatif ; il permet d'indiquer que le mot précédent se réfère à une personne. Le verbe « travailler », qui dérive de la même racine, s'écrit quant à lui  ; le dernier signe est le déterminatif des « actions physiques ou violentes ».

Il faut insister sur le fait qu'un même signe peut avoir plusieurs fonctions. Soit le signe  (O1), qui représente un plan de bâtiment ; il peut fonctionner comme :

- logogramme dans  « *pr* », « *maison* » ;
- phonogramme dans  « *prj* », « *sortir* » ;
- déterminatif/classificateur dans  « *mnqb* », « *pièce fraîche* ».

On peut même le rencontrer plusieurs fois dans un mot, avec des fonctions distinctes : dans  « *pry.t* » « groupe d'immeubles », la première occurrence du signe est un phonogramme ; la dernière est un déterminatif.

En outre, les logogrammes et les déterminatifs peuvent être rassemblés dans une catégorie commune, celle des sémogrammes (*e.g.* Lincke et Kammerzell 2012, 59), c'est-à-dire la classe des signes qui possèdent, dans un emploi donné, un signifié. Ainsi, quand le signe du bâtiment O1 évoqué ci-dessus est employé comme logogramme ou comme déterminatif, il possède toujours un sens : « maison » dans le premier cas et « appartenant à la classe des bâtiments » dans le second. En revanche lorsqu'il est employé comme phonogramme avec la valeur *pr*, aucun signifié ne lui est associé, c'est un simple signifiant.

Le principe de codage que nous formulons repose sur cette distinction binaire :

- 1) Les variantes des signes employés comme phonogrammes ne sont pas susceptibles de modifier le sens d'une occurrence en contexte. Il en découle qu'elles ne doivent pas (nécessairement) être codées dans un corpus annoté.
- 2) Les variantes des signes employés comme sémogrammes, parce que ces derniers possèdent un signifié et sont iconiquement liés au référent, sont susceptibles de modifier le sens d'une occurrence en contexte et doivent être codées dans un corpus annoté lorsqu'elles sont significatives.

---

5. Selon la terminologie proposée par Goldwasser (voir *e.g.* 2002).

#### 4.1.1. Les phonogrammes

Illustrons le principe concernant les phonogrammes avec un signe qui connaît de très nombreuses variantes dans ses emplois comme phonogramme avec la valeur *šsp*, le signe  (O42), qui représente une barrière entourant un naos primitif.

Le *manuel de codage* (Buurman *et al.* 1988, 162) propose plus de quinze formes différentes pour ce signe sous les codes O42, O42A et O43 qui ont toutes reçu des codes O42A, O42B, etc. (e.g. ) dans les extensions plus récente des listes de signes hiéroglyphiques. Comme on l'observe, le nombre et l'orientation des piquets varient, aussi bien que leurs formes et points d'ancrage. Cela conduit les encodeurs à être aussi précis que possible dans le choix de la forme du signe *šsp* lors du codage. Cependant, notre analyse du corpus a montré que les choix étaient la plupart du temps une grossière approximation et n'apportaient pas d'informations propres à faciliter une enquête paléographique ultérieure. Cela ne découle pas seulement de mauvais choix de variantes de la part du codeur, mais également du fait que les quinze variantes disponibles sont loin de couvrir le champ des formes attestées pour ce signe. En ce sens, la prise en compte des variantes de ce phonogramme génère à l'heure actuelle du bruit inutile dans la constitution du corpus et le signe standard O42 pourrait être, sans perte d'information et de manière plus cohérente, employé dans tous les cas.

Nous pensons que ces observations concernant le signe O42 et ses variantes peuvent être généralisées à l'ensemble de la classe des signes employés comme phonogrammes : il faut favoriser dans ce cas une approche par « caractère ». Si toutefois l'on souhaitait tenir compte des variantes, encore faudrait-il que ces dernières soient explicitement déclarées comme telles afin de ne pas rendre problématiques des recherches à venir sur un caractère particulier.

#### 4.1.2. Les sémogrammes

Le cas des signes employés comme sémogrammes est différent. En effet, le hiéroglyphe entretient alors un lien iconique avec son référent, qu'il s'agisse d'une représentation directe, ou d'un lien métonymique. Des modifications du signe, en particulier lorsqu'il est déterminatif, peuvent donc, soit préciser le sens du mot (l'idéogramme du mot « chef », qui désigne généralement un chef étranger, a souvent, dans les versions soignées, des traits caractéristiques qui indiquent l'ethnie du chef en question), soit même proposer une deuxième lecture, graphique, du texte. Nous avons rencontré un exemple illustrant parfaitement ce point dans un texte gravé dans le temple de Kanaïs.

Ce temple se situe dans le désert, à l'est de la ville d'Edfou, sur le chemin des mines d'or. Le roi Séthy 1<sup>er</sup> avait fait creuser là un puits. Dans un texte gravé dans le temple, le nom du roi est suivi d'une épithète assez banale « le bon berger qui fait vivre ses troupes » (Schott 1961, pl. 19, texte A, col. 2). Le texte, tel que codé dans Ramsès, écrivait le mot berger avec le logogramme  (A344), qui représente plutôt un porteur d'eau, au lieu du signe  (A33) attendu normalement. En

retournant à l'original, on s'aperçoit que le signe représente un berger, avec son baluchon, mais aussi une cruche d'eau : . A priori, ce signe n'est pas répertorié<sup>6</sup>. Le point intéressant ici est que le choix du logogramme permet au scribe, tout en écrivant le mot « berger », de faire allusion au contexte précis de l'inscription : le roi ici, est un « bon berger qui fait vivre ses troupes », justement parce qu'il leur fournit de l'eau.

Ainsi que l'illustre l'exemple qui précède, les variations des sémogrames sont très souvent significatives et, par bonheur, les éditeurs sont souvent sensibles à ce phénomène. Ils ont alors à cœur, s'ils le peuvent, de reproduire le mieux possible le signe observé. Il nous faut tenter de respecter leurs choix dans le codage en corpus.

#### 4.2. Le système d'écriture

Encore ce principe n'est-il valide que pour les éditions de textes originellement écrits en hiéroglyphes. En effet, lorsque ces variations dans la graphie des sémogrames sont observées dans l'édition de textes hiératiques, elles ne sont généralement – comme on l'a vu ci-dessus (voir section 3.2.2) – que le fait de choix opérés par les éditeurs dans leur standardisation hiéroglyphique d'un original hiératique éminemment plus cursif. Les variantes graphiques devraient donc, sauf exceptions<sup>7</sup>, être dans ce cas systématiquement ignorées et le codage demeurer aussi générique que possible en favorisant une approche par *caractère*.

### 5. Conclusion : pour une refonte de la liste des signes hiéroglyphiques

Sur base de l'ensemble des observations qui précèdent, nous pensons raisonnable d'avancer, en guise de conclusions, quelques propositions en vue de la reconstruction progressive et motivée d'un répertoire hiéroglyphique<sup>8</sup>.

---

6. En cours de rédaction, nous nous sommes aperçus que le signe A419 ressemblait fortement à notre hiéroglyphe : . En agrandissant le signe, l'objet circulaire porté par A419 semble être une version du phonogramme  (Aa1), translittéré *h*, mais le détail est très petit, et il est difficile de savoir si cette ressemblance est voulue. L'absence de documentation rend donc extrêmement difficile la tâche du codeur, comme celle du dessinateur de polices hiéroglyphiques.

7. Le répertoire des signes hiératiques est essentiellement assez restreint, aux alentours de 800 signes (Möller, 1927). Il arrive que le scribe étende ce répertoire, mais les graphies qu'il emploie alors sont bien plus proches de l'*hiéroglyphe cursif* que du vrai hiératique, et sont immédiatement repérables. Voir en ce sens les remarques de Meeks (2010, 295-296) pour les différences entre paléographies hiéroglyphique et hiératique.

8. On se référera en particulier aux propositions autorisées qui furent avancées par Schenkel (1974, *in tempore non suspecto*).

### 5.1. Un répertoire documenté

Le problème, essentiellement pour les sémogrammes, est souvent de repérer le bon signe dans le répertoire existant, s'il s'y trouve, et, dans le cas contraire, de pouvoir l'y ajouter. Pour permettre cette identification, les signes doivent être correctement documentés, ce qui n'est actuellement pas le cas. Il faut donc construire un répertoire hiéroglyphique suffisamment organisé pour qu'il soit possible d'y retrouver un signe lors du codage d'un texte.

Ce répertoire doit en outre être évolutif, afin de pouvoir y intégrer de « nouveaux » signes, mais, bien évidemment, assez contrôlé pour que son contenu soit raisonnablement fiable. Le processus exact de construction dudit répertoire est probablement un problème ouvert dans l'absolu, mais il est raisonnable de l'envisager dans le cadre fermé des participants à un projet comme Ramsès, voire à plusieurs projets collaborant ensemble à la construction de corpus hiéroglyphiques.

#### 5.1.1. Description des signes

La description formelle de chaque signe doit comporter les emplois répertoriés de celui-ci (comme phonogramme, logogramme, etc.). Outre son intérêt propre, la valeur d'un signe est souvent une manière rapide d'accéder à celui-ci pour le codeur. Bien entendu, l'ouverture du répertoire permet d'ajouter de nouvelles valeurs à un signe déjà documenté.

En outre, les traits distinctifs du signe doivent être décrits, de préférence en utilisant un vocabulaire contrôlé. Dans cette optique, il est intéressant de remarquer que de nombreux hiéroglyphes complexes peuvent être décrits en termes d'autres signes plus simples. Ainsi, le signe  (A10) peut-il être décrit comme le signe  (A1) tenant la rame  (P8).

Des tentatives de description formelle de l'iconographie égyptienne en général ont été réalisées dès les années 1980, et l'idée a été avancée (van der Plas et Vergnienx, 1994) d'utiliser comme unité de description iconographique les hiéroglyphes eux-mêmes. Il faut sans doute nuancer cette position. D'une part, il existe des éléments d'hiéroglyphes qui ne sont pas eux-mêmes des hiéroglyphes. Par ailleurs, le vocabulaire hiéroglyphique peut manquer de termes suffisamment généraux ; il comporte par exemple une multitude de vases. Mais, par là même, aucun de ces signes ne peut servir à désigner, de manière générique, un « vase ».

Outre un vocabulaire de description des éléments des signes, il serait nécessaire de disposer d'un ensemble de prédicats afin de décrire les relations entre les signes en question, comme « coiffé de », « tenant dans main droite », etc.

La description doit enfin être complétée par des références, bibliographiques et iconographiques, aux sources originales (photographies, fac-similés).

### 5.1.2. Séparation entre code et description du signe

Dans la représentation traditionnelle, le code est vu comme une description sommaire du signe, car il lui assigne une catégorie taxonomique. Nous pensons qu'il faut considérer les codes comme arbitraires, et les compléter par une description explicite (section 5.1.1). Une telle approche permettra de résoudre un certain nombre de problèmes.

Tout d'abord, dans la classification traditionnelle, chaque signe est identifié par rapport à son référent (A pour homme, O pour bâtiments, etc.). Cependant, un signe peut être individualisé sans que soit identifié son référent, ou même avoir plusieurs référents possibles. Ainsi, le signe , N18, représente, selon les contextes, une île ou un pain. Ensuite, un signe composite peut relever de plusieurs classes : le signe A254 de la figure 6 doit-il être assigné à la catégorie A (humains), comme son code l'indique, ou E (mammifères) ?



Figure 6. Le signe hiéroglyphique A254

Enfin, les frontières des catégories ne sont pas étanches. Dans l'introduction de la collection « paléographie hiéroglyphique », Meeks propose en ce sens un nouvel ensemble de catégories, qui permet d'éviter un certain nombre d'incohérences dues aux catégories traditionnelles. Il y redéfinit, entre autres, la catégorie W comme « récipients en toutes matières », en précisant que *les récipients représentés contenant des céréales, des fruits (...) sont à classer ici et non dans les autres catégories* (Meeks, 2005, XXII). Cependant, dans le volume 2 de la même collection (Haring, 2006, 116), le signe du boisseau est classé dans la catégorie U, « agriculture, artisanat, métiers et outillage ». Cela ne signifie pas qu'il n'existe pas de classements plus ou moins pertinents. Néanmoins, on y rencontrera toujours des cas limites. Dissocier l'identité d'un signe de son classement résout le problème en permettant d'associer plusieurs classes descriptives au même hiéroglyphe. On passe d'une taxinomie arborescente à une représentation en réseau, à la fois moins contraignante et plus riche.

Ensuite, le code étant arbitraire et fixe, la description peut être affinée, voire modifiée sans conséquence sur les textes déjà codés<sup>9</sup>. Il est alors plus facile d'envisager la construction progressive de la liste des signes, *en parallèle et non en préalable* avec une activité de codage des textes. Les difficultés rencontrées lors de

9. L'avantage des identifiants sans sémantique est, cela va sans dire, bien connu des concepteurs de bases de données.

la recherche d'un signe donné peuvent même être mises à profit pour enrichir ou modifier la description de ce signe.

### 5.1.3. Limitations inhérentes à la nature graphique de l'écriture

Une remarque s'impose cependant : pour le codage des textes monumentaux, l'éventail des choix du codeur est extrêmement étendu. Le répertoire utilisé dépend alors beaucoup des problématiques envisagées par les concepteurs des bases. Il est aussi nécessaire de penser à la fiabilité des informations codées. Représenter un texte avec une précision extrême peut être trompeur, si le reste de la base est codé dans une optique différente. De fait, notre expérience montre que la fiabilité du codage dépend, non seulement du codeur et du texte, mais aussi du signe lui-même. Une mutualisation des listes de signes est cependant possible, en considérant qu'il s'agit d'outils. La documentation associée à chaque signe devrait permettre de faire des choix pertinents.

## 5.2. Un répertoire structuré

En plus d'être documenté et décrit aussi finement que possible, le futur répertoire de signes se doit d'être *structuré*. On entend par là que les relations entre les signes ainsi qu'entre les variantes de tous ordres doivent être déclarées de manière *explicite* et *systématique* : chaque entrée du répertoire devrait en ce sens posséder (au moins) un attribut précisant le degré d'abstraction envisagé pour cette entrée, c'est-à-dire son statut dans la hiérarchie des variantes.

Nous définirons notre codage de manière ensembliste, en considérant un code comme défini par un ensemble des signes attestés dans des documents et partageant des caractéristiques communes. Dans ce codage, nous proposons de distinguer trois niveaux (dans le continuum qui va du type, général et abstrait, à l'occurrence singulière) :

1) Un *graphème*<sup>10</sup> est vu comme une famille de signes. Il est identifié sur base de commutations paradigmatiques et syntagmatiques au sein des lexèmes. Ainsi, si deux formes graphiques distinctes apparaissent dans un même mot au même endroit avec la même fonction, elles sont vraisemblablement des manifestations d'un même graphème. Le graphème possède par conséquent une ou plusieurs fonctions stables (comme phonogramme, logogramme ou classificateur) et sera décrit au niveau iconique de manière suffisamment générale pour être en adéquation avec toutes les occurrences et variantes possibles du signe. La représentation graphique qui est associée à ce graphème ne doit pas nécessairement correspondre à une instance particulière du graphème, puisque aussi bien il s'agit d'un type, mais doit prioritairement être conforme à la description minimale.

2) La *classe* est définie comme sous-ensemble du graphème. Ce sous-ensemble se distingue par des traits iconiques communs et correspond à un degré d'abstraction

---

10. Nous reprenons le terme utilisé dans le *manuel de codage*, avec un sens légèrement différent.

intermédiaire. La classe rassemble donc les variantes d'un graphème qui partagent des traits *iconiques* qui, sans changer le mot, peuvent en moduler le sens : typiquement, l'ajout ou la modification d'un formant dans un signe en modifiera la classe (cf. ci-dessus la spécification ethnique dans le cas de la représentation des chefs étrangers).

3) La *forme* correspond au degré d'abstraction minimal. Elle est subordonnée à la classe : ici entrent en jeu des critères de singularisation d'ordre *stylistique*. Ainsi, au sein d'une même classe, on peut vouloir différencier plusieurs formes en fonction de variations qui ne concernent pas les formants de l'image, mais leur disposition (e.g. un bras plus ou moins penché, le degré d'inclinaison d'un objet, etc.).

On soulignera que l'organisation en trois niveaux hiérarchiques qui est proposée s'adapte bien plus naturellement à une écriture référentielle comme le système hiéroglyphique que la bipartition entre caractères et glyphes d'Unicode, la catégorie *classe* nous permettant d'assurer la médiation entre forme et graphème au niveau iconique.

## Bibliographie

- Bunz C.-M. (2000). *Encoding Scripts from the Past: Conceptual and Practical Problems and Solutions*. Unicode Technical Note 3 <http://www.unicode.org/notes/tn3/>.
- Buurman J., Schimmelpenninck van der Oije A.H.J. (1976). The Composing of Hieroglyphic Texts by Means of a Computer. *Göttinger Miszellen* 19, 7-15.
- Buurman J., Grimal N., Hainsworth M., Hallof J., van der Plas D. (1988). *Inventaire des signes hiéroglyphiques en vue de leur saisie informatique*. Mémoires de l'Académie des Inscriptions et Belles Lettres, Institut de France, Paris.
- Cauville S. (2001). *Dendara, le fonds hiéroglyphique au temps de Cléopâtre*. Cybèle, Paris.
- Cauville S. (1997). *Dendara, les chapelles osiriennes*. Institut Français d'archéologie orientale, Le Caire.
- Collombert P. (2007). Combien y avait-il de hiéroglyphes ?, *Égypte, Afrique et Orient* 46, 35-48.
- Epigraphic Survey (1932). *Medinet Habu*. Vol. 2, *Later historical records of Ramses III*, Oriental Institute Publications 9, Univ. of Chicago Press, Chicago.
- Erman A., Grapow H. (1926-1963). *Wörterbuch der ägyptischen Sprache*. Akademie Verlag, Berlin.
- Everson M. (2006). *Sources for the encoding of Egyptian Hieroglyphs* <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3182.pdf>
- Gardiner A.H. (1957). *Egyptian Grammar, Being an Introduction to the Study of Hieroglyphs*. Griffith Institute, Oxford.
- Gardiner A.H. (1929). The transcription of New Kingdom hieratic. *Journal of Egyptian Archaeology* 15, 48-55.
- Gasse A. (1990). *Catalogue des ostraca hiératiques littéraires de Deir-el-Médina*. Nos. 1676-1774. Tome IV - Fasc. 1, Institut français d'archéologie orientale, Le Caire.

- Goldwasser O. (2002). *Prophets, lovers and giraffes: wor(l)d classification in ancient Egypt*. Classification and Categorization in Ancient Egypt 3; Göttinger Orientforschungen, 4/38, Harrassowitz, Wiesbaden.
- Gourdon Y. (2013). The AGÉA database project, in Polis S., Winand J., Gillen T. (eds). *Text, Languages & Information Technology in Egyptology*, *Ægyptiaca Leodiensia* 9, Liège, 163-168.
- Gozzoli R. (2013). Hieroglyphic Text Processors, Manuel de Codage, Unicode, and Lexicography, in Polis S., Winand J., Gillen T. (eds). *Text, Languages & Information Technology in Egyptology*, *Ægyptiaca Leodiensia* 9, Liège, 89-101.
- Hallof J., van den Berg H. (1992). Thot – Zum Konzept eiens Programs zur Analyse Ägyptischer Texte. *Informatique et égyptologie* 8, 42-46.
- Haring B. (2006). *The tomb of Sennedjem (TT1) in Deir el-Medina: palaeography*. Institut français d'archéologie orientale, Le Caire.
- Janssen J. (1991). *Late Ramesside Letters and Communications*. British Museum Press, Londres (= Hieratic Papyri in the British Museum, 6).
- Janssen J. (1972). Les listes de signes hiéroglyphiques, in Sauneron S. (ed.), *Textes et langages de l'Égypte pharaonique*, Institut français d'archéologie orientale, Le Caire, 57-66 (reprise d'un article publié dans *Chronique d'Égypte XXVII/53*, 1952, 88-97).
- Kitchen K. A. (1969-1990). *Ramesside Inscriptions*, Oxford.
- Legrain G. (1903). Second rapport sur les travaux exécutés à Karnak du 31 octobre 1901 au 15 mai 1902, *Annales du Service des Antiquités de l'Égypte* 4, 1-40.
- Lincke E.-S., Kammerzell Fr. (2012). Egyptian classifiers at the interface of lexical semantics and pragmatics, in Grossman E., Polis S., Winand J. (eds.), *Lexical Semantics in Ancient Egyptian*, *Lingua Aegyptia Studia Monographica* 9, Widmaier Verlag, Hamburg, 55-112.
- Meeks D. (2010). De quelques 'insectes' égyptiens, entre lexique et paléographie, in Z. Hawass, P. Der Manuelian & R.B. Hussein (eds.). *Perspectives on Ancient Egypt. Studies in Honor of Edward Brovarski*, Le Caire, Supplément aux Annales du Service des Antiquités 40, 2010, 273-304.
- Meeks D. (2007). La paléographie hiéroglyphique. Une discipline nouvelle, *Égypte, Afrique & Orient* 46, 2007, p.3-14.
- Meeks D. (2005). *Les architraves du temple d'Esna : paléographie*, IFAO, Le Caire.
- Möller G. (1927-1936). Hieratische Paläographie: die ägyptische Buchschrift in ihrer Entwicklung von der fünften Dynastie bis zur römischen Kaiserzeit, Hinrichs, Leipzig.
- Nederhof M.-J. (2002). A Revised Encoding Scheme for Hieroglyphs. *Proceedings of the 14th table ronde informatique et égyptologie*, Pise (CD-ROM et <http://mjn.host.cs.standrews.ac.uk/publications/2002d.pdf>).
- Nederhof M.-J. (2013). The Manuel de Codage Encoding of Hieroglyphs Impedes Development of Corpora, in Polis S., Winand J., Gillen T. (eds). *Text, Languages & Information Technology in Egyptology*, *Ægyptiaca Leodiensia* 9, Liège, 103-110.
- van der Plas D., Vergnieux R. (1994). Computer-aided Research on Egyptian Stelae from the Middle Kingdom, *Informatique et Égyptologie*, 9, 139-151.

- Polis St., Rosmorduc S. (à paraître). The functions of hieroglyphic signs: Towards a classification. *Lingua Aegyptia*.
- Polis St., Winand J. (2013). The Ramses project: Methodology and practices in the annotation of Late Egyptian Texts, Hafemann I. (Ed.). *Perspektiven einer corpusbasierten historischen Linguistik und Philologie : Internationale Tagung des Akademienvorhabens Altägyptisches Wörterbuch an der BBAW, 12.-13. Dezember 2011*, Thesaurus Linguae Aegyptiae 4, Berlin, 81-108 (<http://edoc.bbaw.de/volltexte/2013/2433/>).
- Rosmorduc S. (2002). Codage informatique des langues anciennes, le cas des hiéroglyphes égyptiens. *Document numérique* vol. 6/2002, n° 3-4, 211-224.
- Schenkel W. (1974). Gesichtspunkte für die Neugestaltung der Hieroglyphenliste. *Göttinger Miszellen* 14, 31-45.
- Schott S. (1961). *Kanais. Der Tempel Sethos I. im Wâdi Mia*, Vandenhoeck & Ruprecht, Göttingen.
- Sethe K. (1927-1930). *Urkunden der 18. Dynastie*, 4 vol., Akademie-Verlag, Berlin.
- Vernus P. (1990). Les espaces de l'écrit dans l'Égypte pharaonique. *Bulletin de la Société Française d'Égyptologie* 119, 35-56.
- Winand J. (2013). *Les hiéroglyphes égyptiens*. PUF, Paris.