Exemple d'écriture ignorée par Unicode : l'écriture tham du Laos

Grégory Kourilsky

Institut National des Langues et des Civilisations Orientales
2, rue de Lille 75007 Paris
gregory@kourilsky.com

Mots-clés : Unicode, écriture peu dotée, langue-π, écriture non linéaire, signes non connexes, écritures d'Asie du Sud-Est.
Keywords: Unicode, under-resourced scripts, π-language, non linear script, non-connex scripts, South-East Asian Scripts.

Résumé : L'écriture tham, employée pour noter les textes bouddhiques du Laos, est une *écriture* peu dotée informatiquement, alors que les *langues* qu'elle transcrit (lao et pali) ont fait l'objet d'une informatisation avancée. Considérant les raisons sociologiques et techniques de ce délaissement, nous proposons des pistes de réflexion pour y remédier en envisageant notamment un codage Unicode ainsi qu'une méthode de rendu.

Abstract: Tham Script, used to write Buddhist texts in Laos, is an under-resourced *script* although the *languages* that it transcribes (Lao and Pali) are well computerized. Understanding the sociological and technical reasons of this neglect, we'll propose some bases of reflection to mend it, by considering an Unicode Encoding and some rendering methods.

1 Un système d'écriture mal doté informatiquement

L'écriture tham est l'une des deux écritures officielles du Laos. Moins connue que l'écriture lao dont l'usage est courant, le tham est employé exclusivement pour noter les textes bouddhiques. Cette écriture, d'origine indienne et vraisemblablement établie par les Môn aux alentours du XVe siècle (Ferlus, 1995), a la particularité de transcrire deux langues, le pali (langue du bouddhisme *theravāda*) et le lao, avec des règles et certains caractères spécifiques à la notation de chacune de ces deux langues. L'écriture tham est considérée comme sacrée et son emploi est essentiellement réservé aux bonzes (moines bouddhistes) et aux lettrés[1].

[1] Le tham du Laos est très proche des écritures yuon (Nord de la Thaïlande), khün (États Shans de Birmanie) et lü (région du Xishuanbanna de Chine) si bien que l'on qualifie parfois l'ensemble de ces écritures d'« écritures tham », notamment parce que celles-ci sont surtout employées pour noter les textes bouddhiques (le terme thaï-lao *tham* dérive du mot pali *dhamma*, « doctrine bouddhique »). Précisons que les écritures khün, lü et yuon ne sont guère mieux dotées informatiquement que le tham du Laos (on note néanmoins l'existence de plusieurs polices TrueType yuon).

Écriture tham : ᨾᩮᩣᨲᩖᨾᨶᨠᩣᨲᨶᨶᨲᨲᨲᩣᨲᩣᨾᩣᨾᩩᩫᩣᩈᩣ [(2)]

Parmi les deux écritures officielles du Laos, le lao et le tham, seule la première a fait l'objet de travaux informatiques (création de nombreuses polices de caractères, prise en compte par Unicode, traitements de textes, etc.) tandis que la seconde a été presque totalement délaissée. En effet seulement trois polices existent actuellement pour saisir le tham : la *LaoDhamma* (TrueType) créée par S. Morey qui fonctionne sur Macintosh ; la *Vat Sène* (PostScript), élégante et très complète, fonctionne sur Macintosh et a été réalisée par l'École Française d'Extrême-Orient (ÉFEO). Cependant, cette dernière ne paraît stable que sur les logiciels In-design et X-Press et n'a pas été mise à la disposition du public. La troisième est la *ThamStandard* (TrueType) pour PC mise au point par nos soins et utilisée pour la rédaction du présent article. Trois facteurs peuvent expliquer ce caractère anormalement peu doté de l'écriture tham :

- le tham est une écriture peu connue de la population laotienne. Elle est employée exclusivement par les moines bouddhistes à l'intérieur des pagodes,

- le tham est une écriture sacrée, possédant une fonction religieuse, voire magique. Cette dimension implique que cette écriture est plus qu'un système de transcription et son traitement informatique soulève des questions d'ordre socio-religieux,

- le système d'écriture tham est complexe : il transcrit deux langues très différentes (le pali et le lao), fait usage de signes non connexes et de nombreux signes souscrits et suscrits. Cette complexité est accentuée par le fait que cette écriture n'a jamais été l'objet d'une réelle codification, ce qui implique des graphies et des orthographes très variables selon les textes.

C'est de ce dernier facteur dont il sera question ici, les deux premiers étant traités dans (Kourilsky, 2005).

2 Trois difficultés pour une informatisation de l'écriture tham

2.1 Écriture non linéaire à signes non connexes

À l'instar des autres écritures d'origine indienne (devanagari, lao, thaï, khmère, etc.), on remarque en tham une hétérogénéité dans la disposition des signes-voyelles par rapport à la consonne. Exemples avec la consonne ᨠ *k* [3]:

Voyelle inhérente à la consonne : ᨠ /ka/,

Signe-voyelle à la suite de la consonne : ᨠᩣ /kaː/,

Signe-voyelle avant la consonne (signe *antéposé*) : ᩋᨠ /keː/,

Signe-voyelle au-dessus de la consonne : ᨠᩤ /kiː/,

---

[2] « Je salue le Bienheureux, le Vénérable, le Parfaitement éveillé » (Bizot, Lagirarde, 1996).

[3] Les lettres tham isolées et les mots pali sont transcrits conformément à la transcription latine officielle du sanskrit fixée au 10e Congrès des orientalistes (1894), le pali suivant *grosso modo* les mêmes règles de prononciation que le sanskrit. Les mots lao écrits en tham sont transcrits en API et, le cas échéant, en écriture lao.

Exemple d'écriture ignorée par Unicode : l'écriture tham du Laos

Signe-voyelle en-dessous de la consonne : ᨠᩩ /ku/,

Signe-voyelle de part et d'autre de la consonne : ᨠᩮᩬᩴ /kjaʔ/

Le tham fait usage également de consonnes de forme souscrite et suscrite :

À la suite de la consonne : ᨠᩬᩣ /klaː/,

Avant la consonne (signe *antéposé*) : ᨠᩮᩣ /kʀeː/,

Au-dessus de la consonne : ᨠᩤᩩ /kuŋ/,

En-dessous de la consonne : ᨠᩥᩫ /kim/.

La position des signes-voyelles et aussi des signes consonantiques souscrits ou suscrits (dits de forme *subjointe*) incite à nuancer l'affirmation que l'écriture tham s'exécute de gauche à droite (Bizot, 2001) comme le montrent les exemples suivants :

ᨠᩮᩬᩫ *ra b h ma* = « *brahma* » ᨡᩥᨠᩮᩣ *i n or k ō* = « *nikrō* »

2.2 Deux langues, deux systèmes

Le tham est un système d'écriture qui permet de transcrire deux langues, le pali et le lao. Les règles et propriétés du système d'écriture tham varient selon que celui-ci note l'une ou l'autre de ces deux langues. Aussi, il arrive qu'un même mot s'écrive différemment en tham en fonction de la langue transcrite, par exemple pour un mot pali passé dans la langue lao : Ainsi *taṇhā* , « désir » sera noté (1) ᨠᨠᩫᩣ dans un texte pali mais (2) ᨠᩬᩴᩣ [ຕັນຫາ] dans un texte lao.

En notation du pali, la consonne qui suit une consonne dévoyellée prend sa forme subjointe ((1) : ᩫ *h*). En notation du lao, c'est en position finale ou médiane qu'une consonne peut prendre sa forme subjointe ((2) : ᩬ *n*). Pour cette raison, nous avons choisi d'employer la terminologie suivante :

- P-tham : système d'écriture tham pour la notation de la langue pali,
- L-tham : système d'écriture tham pour la notation de la langue lao.

De plus, les graphies de certaines consonnes subjointes diffèrent selon la langue transcrite :

Valeur	Forme nominale	Forme subjointe p-tham	Forme subjointe l-tham
k	ᨠ	ᩫ	ᩫ ou ᩬ
ṅ	ᨡ	ᩬ	ᩬ ou ᩬ
n	ᨢ	ᩬ	ᩬ ou ᩬ
l	ᨣ	ᩫ	ᩬ

Figure 1 : Variations graphiques entre subjointes p-tham et l-tham

2.3 Une écriture non fixée

Comme l'écriture tham n'a jamais fait l'objet de réelle codification, on trouve des mots écrits avec des orthographes différentes selon les textes, si bien qu'« il est pratiquement impossible de trouver deux manuscrits identiques, car chaque nouvelle copie subit des modifications ou des corrections effectuées par un lettré qui estime, à tort ou à raison, être dans son bon droit[4]. » (Peltier, 2000). En fait les copistes privilégient avant tout l'harmonie esthétique, critère subjectif par excellence. Cela est surtout flagrant en l-tham :

ᩋᩬᨾ /nɔ̆ːk/	ᩋᩬᨾ /mɔ̆ːk/	ᩃᩬᨾ /lɔ̆ːk/	ᩋᩬᨾ /nwăːk/	ᩋᩬᨾ /mwăːk/
ᩉᩣᩴ	ᩉᩣᩴ	ᩉᩣᩴ	ᩉᩣᩴ	ᩉᩣᩴ
ᩉᩣᨠ	ᩉᩣᨠ	ᩉᩣᨠ	ᩉᩣᨠ	ᩉᩣᨠ
ᩉᩣ᩠	ᩉᩣ᩠	ᩉᩣ᩠	ᩉᩣ᩠	ᩉᩣ᩠

Figure 2 : Mots à plusieurs orthographes (Phouy, 1943a)

Ainsi, les copistes usent à l'envi d'innovations orthographiques, parfois même de simplifications dans l'écriture des mots qui ont pour objectif de réduire le nombre d'idiographèmes dans certains mots courants. On appelle ces formes simplifiées des *contractions* [ᨾᩣᩉᩮᩢ] /kʰamɲɔ̆/. Ces dernières, contrairement aux mots lao qui suivent normalement l'orthographe phonétique, ne sont pas toujours aisément déchiffrables et il est souvent nécessaire de les connaître par cœur pour déceler de quel mot il s'agit. Il faut également souligner que les contractions peuvent varier d'un texte à l'autre. Exemples :

ᨴᩤᩥ = ᨴᩥᩉᩤᩥ /diːlĭiː/ ; ᨠᩴᩣ = ᨠᩴᩣᩁᩤ /kʰanvàːl/ ; ᨶᩥ = ᨶᩴᩣᩁᩤ /anvàːl/, etc.

L'aspect ultime de ces contractions se trouve dans le phénomène que nous avons appelé *glyphe syllabique*, où un signe graphique unique note une syllabe (autre que le cas consonne + /a/ bref inhérent) : ᩊ = ᩃᩮᩬ /lɛ/ ; ᩐ = ᨶᩥ /rɯː/, ◌ᩮᩬ = ᩁᩮᩬ /áo/, etc.

3 Solutions envisagées pour informatiser le tham

Nous avons en premier lieu réalisé une police de format TrueType que nous avons appelée *ThamStandard*. En tenant compte des signes multi-usages et des signes non connexes que nous avons décidé de considérer comme signes unitaires, nous avons besoin de 132 signes (i.e. *caractères*) accessibles au clavier. La police ThamStandard comporte ces 132 caractères, en plus d'une dizaine de variantes graphiques. Nous avons également conçu un clavier virtuel utilisable à partir d'un clavier AZERTY, nommé *ThamFrance* qui offre un maximum de correspondance phonétique entre les caractères latins et tham (par exemple ᨠ *k* se saisie avec la touche latine [k], ᩣ *ā* avec la touche [A], etc.). Si un tel système a l'avantage de ne nécessiter que la seule installation de la police *ThamStandard* et est donc utilisable sur n'importe quel ordinateur PC ou compatible, l'accès à certains signes manque d'ergonomie. En effet, les touches du clavier sont insuffisantes

[4] On peut aussi penser que le copiste souhaitait bénéficier de l'acquisition de mérites découlant de l'amélioration d'un manuscrit. Ce phénomène s'explique aussi par le fait que la copie des manuscrits est souvent confiée à des novices ou des jeunes moines qui ne maîtrisent que sommairement l'écriture tham.

pour saisir les 132 signes *minima* nécessaires pour représenter l'intégralité des lettres tham, et il faut alors recourir à des combinaisons de touches parfois tortueuses pour afficher certains signes. Beaucoup d'écritures indiennes et d'Asie du Sud-Est (devanagari et khmère en particulier) rencontrent le même problème[5].

Une manière de remédier au caractère peu doté de l'écriture tham tout en apportant une solution aux problèmes non résolus par la police *ThamStandard* et le système de saisie *ThamFrance*, serait d'intégrer le tham au Standard Unicode. Dans son objectif de normaliser les codages des caractères du monde entier afin de permettre une compatibilité entre les différentes polices et plates-formes, quel que soit le pays ou la langue, le Standard Unicode a établi des règles de codage pour l'ensemble des écritures figurant au Standard. Une de ces règles est de ne coder que les *caractères abstraits* (unités fondamentales de codage) et non les *glyphes*[6], ces derniers étant les représentations (parfois sous plusieurs formes) de ces caractères abstraits. Pour résumer, nous dirons qu'à un caractère abstrait correspond un ou plusieurs glyphes, et inversement plusieurs caractères abstraits peuvent être représentés par un glyphe identique. Ce système de codage permet (en théorie) de simplifier considérablement la saisie puisque c'est le système de mise en œuvre (par exemple l'Uniscribe de Windows) qui affichera les glyphes adéquats en fonction des contextes. En examinant le traitement par Unicode des écritures indiennes et dérivées (devanagari, tamoule, thaï, lao, birmane, khmère et tibétaine), qui font toutes usage de signes non connexes et de formes antéposées, souscrites et suscrites, on s'aperçoit que les problèmes ont été traités différemment selon les écritures. En fait on peut distinguer deux modèles : le modèle « indien » et le modèle « thaï-lao » qui divergent dans le codage des voyelles et dans l'ordre de stockage en mémoire[7].

Le modèle « indien » (devanagari, tamoul, khmer, etc.) suit à la lettre les principes d'Unicode en codant les voyelles en tant que phonèmes, ou reconnus comme tels dans les règles descriptives de ces langues. C'est ainsi que les voyelles composées de plusieurs idiographèmes (et quelle que soit leur position par rapport au signe consonantique) sont codées indépendamment, chacune d'elles représentant un *caractère unique*. Ainsi la voyelle tibétaine ི /i/ est-elle codée U+0F73, la voyelle khmère ើ /ɯə/ est codée U+0BCA, etc. Ce modèle respecte les principes d'Unicode mais présente des difficultés tant de mise en œuvre que d'utilisation (le khmer en particulier). D'autre part ce modèle code en mémoire les caractères vocaliques après la consonne, y compris les signes antéposés. Les consonnes subjointes antéposées sont stockées également après la consonne qui la précède en lecture.

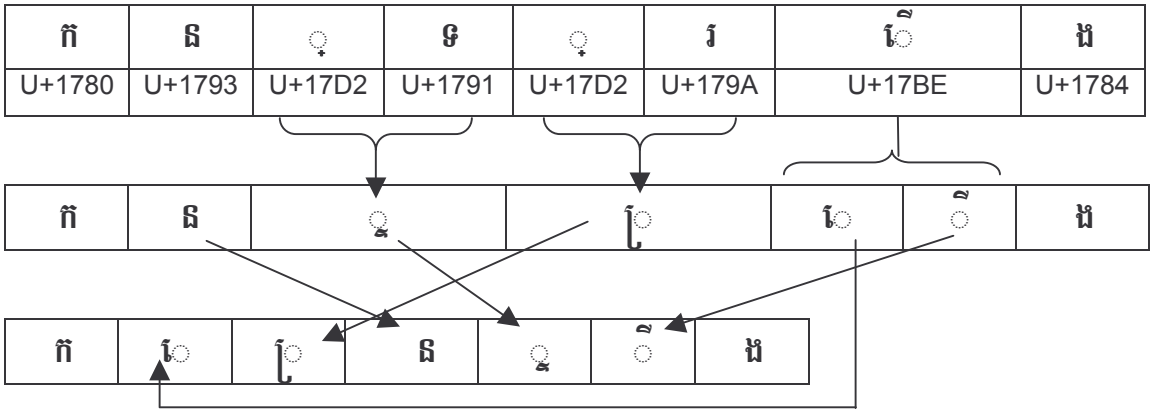
Exemple (d'après Bauhanh, 2002) : saisie du mot khmer កន្ត្រើង /kɑːntrəːŋ/

---

[5] La police khmère *Kdol*, mise au point par l'ÉFEO et qui intègre un grand nombre de ligatures respectant les graphies de l'écriture khmère employées dans les manuscrits des XVIIe-XIXe siècle, comporte ainsi 438 signes (Bizot, 1992).

[6] Le terme *glyphe*, désignant étymologiquement un signe gravé, était traditionnellement employé pour désigner les caractères maya avant d'être repris par les concepteurs d'Unicode.

[7] Nous assimilerons l'ordre logique (le stockage en mémoire) à l'ordre de saisie : « Les textes Unicode (...) sont stockés en mémoire en ordre logique. Cet ordre correspond *grosso modo* à l'ordre dans lequel le texte est saisi au clavier » (Andries, 2002).



On relève la complexité de l'ensemble des opérations que le système doit réaliser[8] :

- transformation des consonnes de forme nominale ◌ រ et ◌ ◌ en forme souscrite ◌ ◌ [9] et ◌ ◌,
- repositionnement du glyphe ◌ ◌,
- décomposition de ◌ ◌ ◌ en ◌ ◌ et ◌ ◌,
- repositionnement de ◌ ◌ (juste avant ◌ ◌).

On pourra notamment remarquer l'opération pour le moins déroutante qui consiste à devoir décomposer ◌ ◌ ◌ en ◌ ◌ et ◌ ◌ , alors même qu'Unicode insiste pour coder le caractère U+17BE ◌ ◌ ◌ séparément. D'une manière générale, on ne peut que constater le manque de souplesse du mode de saisie des polices khmères Unicode qui est souvent en totale contradiction avec le mode d'écriture manuscrite. Pour cette raison, Unicode précise que le système de saisie du khmer Unicode n'est pas intuitive et doit faire l'objet d'une formation (Unicode 4.0, p. 277). Les arguments avancés par certains experts pour cet encodage « logique » sont de faciliter le tri, la recherche lexicographique et la reconnaissance vocale. On peut cependant douter de la pertinence de ces priorités, dont l'une des contreparties est la nécessité d'un apprentissage complexe, au regard d'une population en lutte contre un illettrisme conséquent causé par les évènements tragiques de ces dernières décennies.

À l'inverse, le modèle « thaï-lao », mis sur pied avant l'intégration de ces écritures dans Unicode, suit un système local (la norme thaï TIS 620) qui code les idiographèmes vocaliques et non les signes-voyelles dans leur intégralité. Les caractères vocaliques sont donc des *signes visibles* et non plus obligatoirement des *phonèmes*. Les voyelles composées de plusieurs idiographèmes sont alors rendues par l'intermédiaire de plusieurs *caractères*. D'autre part, avec ce modèle, on saisit les signes dans un ordre purement graphique. Ce système déroge aux principes d'Unicode, mais offre des avantages pratiques. Exemple :

0EC0 ◌ + 0E9A ◌ + 0EB7 ◌ + 0EAD ◌ → ◌ ◌ ◌ /bɯaː/ (avec ◌ /b/ et ◌ ◌ ◌ /ɯːa/)

On comprend que le codage des caractères lao a été pensé suivant une logique graphémique, puisque le caractère U+0EAD ◌ LAO LETTER O est employé aussi bien pour saisir la consonne ◌ /ʔ/

---

[8] Opérations auxquelles peuvent s'ajouter le rejet des saisies incorrectes, la correction automatique, etc.

[9] Voir *infra*, 3.2 pour le rendu des formes subjointes khmères. Notons que dans ce contexte, la consonne ◌ /r/ souscrite prend la forme ◌ ◌, à hampe plus allongée que le glyphe ◌ ◌ habituellement employé.

que l'idiographème vocalique ᩉ qui sert à écrire les voyelles ◌ᩉ◌ /ɔː/ interconsonantique, ᨠ◌̊ᩉ /ɯa/ et ᨠ◌̊ᩉ /ɯːa/.

Une réflexion sur une intégration du tham à Unicode permet d'envisager des solutions aux trois difficultés évoquées (*supra*, 2). Nous avons provisoirement attribué une zone de code tham (F000-F07F) dans la zone d'usage privé d'Unicode (E000-F8FF), en reprenant l'affectation proposée par V. Berment (Berment, 2004). Cette zone provisoire tham Unicode figure en fin d'article et ne doit être envisagée que comme situation temporaire en attendant l'intégration en bonne et due forme d'une zone tham définitive en accord avec le Standard Unicode[10].

3.1 Écriture non linéaire à signes non connexes

Pour le traitement des voyelles non connexes (*supra*, 2.1), nous avons donc à notre disposition un « modèle indien » qui code les voyelles en tant qu'unité phonologique, et un modèle « thaï-lao » qui code non les voyelles mais les *idiographèmes vocaliques* permettant de les composer graphiquement. Pour le principe de codage des voyelles tham, le « modèle thaï-lao » a été choisi en défaveur du « modèle indien », et ce pour deux raisons :

- le système adopté par le khmer, qui code effectivement toutes les voyelles, y compris les voyelles non connexes, nous a semblé trop complexe à mettre en œuvre et à utiliser, d'autant que l'écriture tham ne bénéficiera pas des moyens humains et financiers mis à disposition pour mettre en place le « khmer Unicode »,
- le système d'écriture l-tham est proche du système d'écriture lao (les deux systèmes notent la même langue), et les graphies des voyelles tham (tant p-tham que l-tham) sont pour la plupart similaires à celles de l'écriture lao. Avec un système de codage « thaï-lao », les utilisateurs familiarisés avec le lao Unicode passeront sans difficulté au « tham Unicode ».

Par conséquent, le mot ᨣᩢ᩵ᩖ᩠ᩖ /kjaʔ/ (avec ᨠ /k/ et ᨣᩣ᩵ᩖ᩠ᩖ /jaʔ/) sera saisi :

F03A ᨣ + F000 ᨠ + F048 ◌᩵ + F041 ◌᩠ + F03B ◌ᩖ → ᨣᩢ᩵ᩖ᩠ᩖ /kjaʔ/

et non : F000 ᨠ + F0XX ᨣᩣ᩵ᩖ᩠ᩖ → ᨣᩢ᩵ᩖ᩠ᩖ /kjaʔ/

3.2 Deux caractères de rendu pour deux systèmes d'écriture

En p-tham, nous pouvons appliquer un système relativement simple pour le rendu automatique des consonnes subjointes. En effet, « la consonne au niveau de la ligne d'écriture est la finale de la consonne qui la précède (*i.e.* dévoyellée), tandis que la consonne placée au niveau inférieur est la consonne liée à la voyelle [qui suit] » (Phouy, 1943b). Si nous reprenons l'exemple du mot ᨶᨦ᩶ᩉᩫ *taṇhā*, la présence du ᩉ /h/ souscrit indique que la consonne ᨦ /n/ qui le précède est dévoyellée. On lira alors /tanhaː/ et non /tanahaː/. Ainsi, il suffit de considérer la présence *théorique* d'un caractère équivalent au *virama* indien qui, associé à une consonne, indique que celle-ci perd sa voyelle inhérente. Le khmer Unicode use d'un tel caractère particulier pour rendre les souscrites, appelé caractère *coeng* (U+17D2), bien que celui-ci ne fasse pas partie de l'écriture khmère :

[10] Une proposition de codage tham sera soumise au Comité technique d'Unicode (UTC) courant 2005.

1785 ᨠ + 17D2 ᩲ + 179A ᨿ → ᨠᩲᨿ /crɑː/

Ce signe *coeng* fonctionne à la manière d'un *virama* en « tuant » la voyelle inhérente de la consonne précédente et indiquant que la suivante est subjointe. La règle p-tham de souscription des consonnes étant similaire à celle du khmer, nous pouvons procéder de même et considérer un *virama* « théorique », sans châsse et sans forme visible qui, associé à une consonne, affiche sa forme subjointe. Nous avons appelé ce caractère *VIRAMA THÉORIQUE THAM* (VTT). Exemple :

F00F ᨲ + F000E ᨶ + F042 VTT + F01E ᨸ → ᨲᨶᨸ *taṇhā*

On notera le cas exceptionnel[11] du ᨿ/ʀ/ antéposé, qui sera saisi dans l'ordre graphique. Dans ce cas la logique linguistique est invalidée, le VTT n'indiquant pas que la consonne ᨷ est dévoyellée :

F042 VTT + F01A ᨿ + F016 ᨷ + F01E ᨸ + F042 VTT + F018 ᨾ → ᨿᨷᨸ *brahma*

Pour la saisie du l-tham, nous pouvons conserver cette même méthode de rendu (avec le caractère F042 VIRAMA THÉORIQUE THAM) et considérer un outil supplémentaire de rendu de glyphes particuliers, sans se soucier de légitimité linguistique. On imagine ainsi un deuxième caractère de rendu similaire au F042 VIRAMA THÉORIQUE THAM. Ce caractère, que nous appellerons SIGNE DE RENDU THAM, codé F043 à la suite du *virama* « théorique », permettra d'afficher les variantes graphiques de certains caractères. De cette manière, selon le glyphe qu'il souhaite voir afficher, l'utilisateur choisira soit le F042 VIRAMA THÉORIQUE THAM, soit le F043 SIGNE DE RENDU THAM.

+	F000 ᨶ	F004 ᨦ	F01B ᨧ	F01D ᨾ
F042 VTT	ᩲᨶ	ᩲᨦ	ᩲᨧ	ᩲᨾ
F043 SRT	ᩲᨶ	ᩲᨦ	ᩲᨧ	ᩲᨾ

Figure 3 : Rendu des variantes graphiques des formes subjointes

Exemples : F00F ᨲ + F038 ᩲ + F042 VTT + F004 ᨦ → ᨲᩲᨦ /tuŋ/

F00F ᨲ + F048 ᩲᨶ + F043 SRT + F004 ᨦ → ᨲᩲᨦ /taŋ/

[11] D'autres cas particuliers sont à traiter mais ne seront pas pris en considération ici : le ᩲᨦ *ṅ* final, les ligatures, etc. Ces signes sont également rendus par l'intermédiaire du caractère F042 VTT (Kourilsky, 2005).

3.3 Affichage des glyphes non contextuels

Il reste à résoudre la troisième difficulté évoquée, à savoir l'absence de codification de l'écriture tham. Un utilisateur doit être en mesure d'afficher les glyphes non contextuels (*i.e.* les formes graphiques non imposées par le contexte, et dépendantes du goût du copiste). La méthode d'utilisation de deux caractères de rendu décrite ci-dessus permet de résoudre en partie ce problème. Les mots l-tham pouvant s'écrire avec plusieurs orthographes possibles comme /mɔ̃ːk/ (*supra*, 2.3) seront rendus soit avec le F042 VIRAMA THEORIQUE THAM (rendu ᩲᩮᩴ), soit avec le F043 SIGNE DE RENDU THAM (rendu ᩲᩮᩴ)[12].

D'autre part, le caractère F043 SRT peut être utilisé pour noter des variantes graphiques tham, indépendamment de la langue transcrite (lao ou pali). Avec ce système, tout caractère a, dans une même police, *potentiellement* deux variantes graphiques pour ses subjointes. Il pourra être intéressant pour un utilisateur souhaitant respecter au mieux les graphies originales d'un texte de disposer d'un éventail de choix dans une police unique. Nous donnons ci-dessous des exemples de saisie de différents glyphes rendus avec les caractères F042 VTT et F043 SRT.

+	F00C ᩮ	F018 ᩯ	F012 ᩰ	F01C ᩱ
F042 VTT	ᩮ᩠ᩮ	ᩮ᩠ᩯ	ᩮ᩠ᩰ	ᩮ᩠ᩱ
F043 SRT	ᩮ᩠ᩮ	ᩮ᩠ᩯ	ᩮ᩠ᩰ	ᩮ᩠ᩱ

Figure 3 : Rendu des variantes graphiques des consonnes subjointes

Quant aux contractions, elles ne poseront pas de problème de rendu puisque celles-ci ne font usage que de signes déjà employés dans l'alphabet (idiographèmes vocaliques ou consonnes subjointes).

Exemple : F00C ᩮ + F037 ᩤ + $\underbrace{\text{F043 SRT + F01B ᩬ}}_{\text{ᩬ᩠ᩬ}}$ → ᩮᩤᩬ᩠ᩬ /diːlĭː/

Enfin, le problème posé par les *glyphes syllabiques* (*supra*, 2.3) peut être facilement résolu, dans la mesure où nous avons opté pour un codage « graphique » suivant le modèle « thaï-lao » (*supra*, 3.1). Ainsi, si l'on veut pouvoir disposer librement (et simplement) des signes ᩬ /lɛ/ et ᩭ /rɯː/, il suffit de leur attribuer un code : F056 ᩬ SYLLABE THAM LEI et F057 ᩭ SYLLABE THAM RYY. Les signes ᩤ /áo/ (contraction de ᩤᩣ) et ᩤ /ɛː/ (variante graphique de ᩤ) seront également codés, respectivement F03B ᩤ SIGNE VOCALIQUE THAM LAO AO et F040 ᩤ SIGNE VOCALIQUE THAM LAO MAE EI.

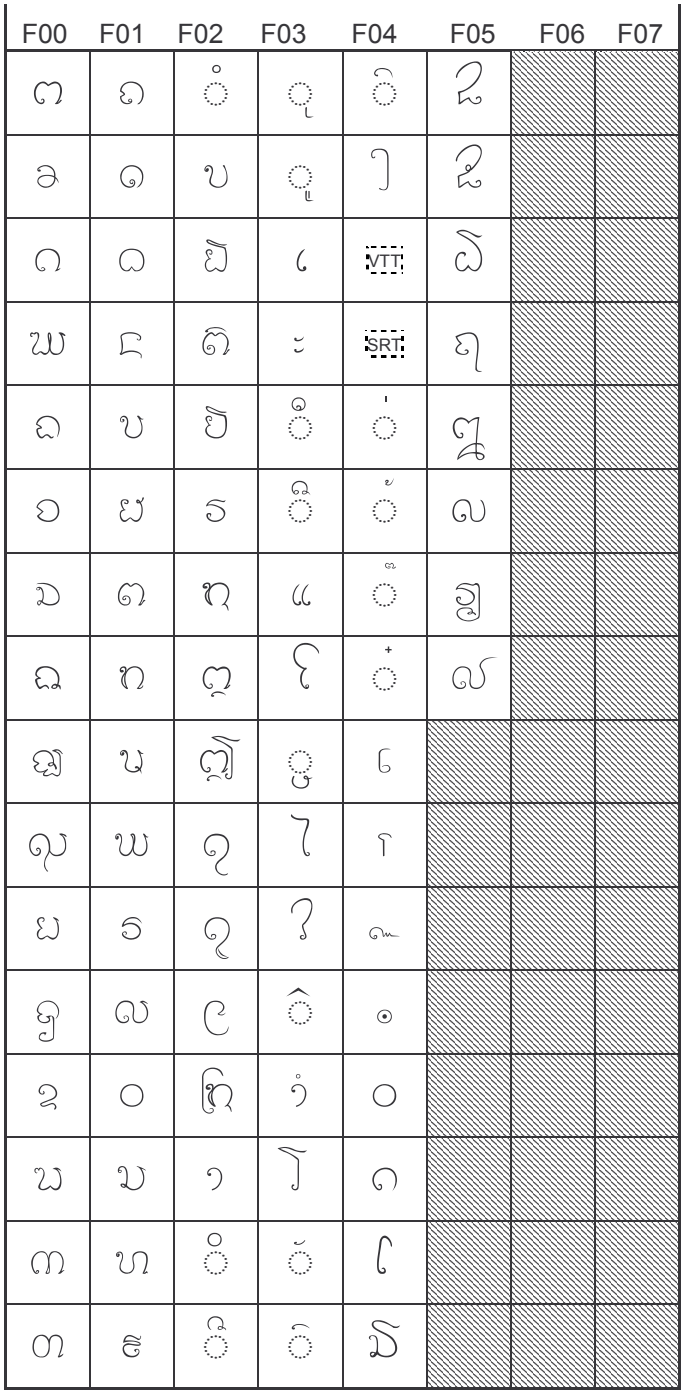
---

[12] Parmi les orthographes mentionnées en 2.3., seule la troisième, plus rare, ne pourra être rendue.

Grégory Kourilsky

Proposition pour une zone tham Unicode

F000 **Tham** **F07F**



Conclusion

Le standard Unicode et les méthodes de rendu proposées par le consortium permettent de résoudre les difficultés inhérentes au traitement informatique des écritures indiennes et dérivées (nombre très important de signes, signes antéposés et non connexes, variantes graphiques d'une même lettre, etc.). Cependant certains partis pris, assez tentant au premier abord, de « logique linguistique », impliquent une mise en œuvre complexe. Nous avons souligné la complexité du système de rendu de l'écriture khmère en raison du stockage en ordre « logique », ce qui a retardé considérablement et pénalise encore l'utilisation du khmer Unicode. Il est permis de penser qu'une intégration de l'écriture tham au standard pourra pallier le caractère informatiquement peu doté de ce système d'écriture. Encore faut-il réaliser cette intégration par un codage et une méthode de rendu simples, intuitifs et accessibles à ceux, éditeurs, étudiants, chercheurs ou encore moines bouddhistes, qui en feront usage.

Références

ANDRIES P. (2000), *Introduction à Unicode et à l'ISO 10646,* Québec, Document numérique, Volume 6, n° 3-4, 51-88.

BAUHANH M. (2002), « Rendering the World's Complex Scripts: A Case Study in Khmer », 21th International Unicode Conference, Dublin, 20 p.

BERMENT V. (2004), *Méthodes pour informatiser des langues et des groupes de langues « peu dotés »*, Grenoble, Thèse de doctorat (Université Joseph Fourier), 277 p.

BIZOT F. (1992), *Le Chemin de Laṅkā*, Paris, ÉFEO, TBC i, Paris, 352 p.

BIZOT F., BIZOT C. (2001), « Écritures bouddhiques d'Asie du Sud-est », *in* :*Histoire de l'écriture, de l'idéogramme au multimédia* (dir. A-M. Christin), Flammarion, 149-153.

BIZOT F., LAGIRARDE F. (1996), ບຄຸດິບດ, *La pureté par les mots*, Paris, Phnom Penh, Chiang Mai, Vientiane, ÉFEO, Textes bouddhiques du Laos, 275 p.

FERLUS M. (1995), « Les circonstances de l'introduction de l'alphabet tham lanna », *in* : *La Thaïlande des débuts de son histoire jusqu'au XVe siècle*, Premier Symposium Franco-Thai, 18-24 juillet 1988, Université Silpakorn, 101-109.

KOURILSKY G. (2005), *Éléments pour un traitement informatique de l'écriture tham du Laos*, mémoire de DREA (Institut des Langues et des Civilisations Orientales), 330 p. (à paraître).

PELTIER, A.R (2000), ຂໍ້ຜົ້ທຽນ, *La fille aux cheveux parfumés*, Vientiane, IRC, 490 p.

PHOUY P. (1943a), *ແບບຮຽນໄວຮຽນອ່ານໜັງສືທັມ ຂຽນເປັນພາສາລາວ [Apprendre rapidement à lire les caractères tham dans les textes lao]*, Bangkok, Kramol Tirannasur, 80 p.

PHOUY P. (1943b), *ແບບຮຽນໄວຮຽນອ່ານໜັງສືທັມ ຂຽນເປັນພາສາບາລີ [Apprendre rapidement à lire les caractères tham dans les textes pali]*, Bangkok, Kramol Tirannasur, 51 p.

UNICODE (2004), *The Unicode Standard 4.0*, http://www.unicode.org.