

Polices de caractères et codage pour le laotien

Pour les polices de caractères du laotien, il faut tenir compte à la fois

- de la graphie des caractères
- du codage des caractères

Tout ce qui est informatisé est codé *numériquement*. Le plus petit élément de codage est le *bit*, qui peut prendre les deux valeurs 0 et 1. L'unité suivante est l'*octet* (en anglais *byte*), qui est un « mot » de 8 bits, ce qui représente 256 possibilités (soit un codage à 256 valeurs, numérotées de 0 à 255 en utilisant la numération décimale à laquelle nous sommes habitués).

00000000 =	0
00000001 =	1
00000010 =	2
00000011 =	3
00000100 =	4
...	
01111111 =	127
10000000 =	128
...	
11111110 =	254
11111111 =	255

Codage Windows

Le **codage de Windows** pour les langues latines se compose d'une table de 256 codes, correspondant aux 256 valeurs possibles d'un octet :

- 0-31 : codes spéciaux (dont le tabulateur, le code de fin de ligne, etc.)
- 32-127 : caractères alphabétiques non accentuées, numériques et ponctuation standard
- 128-255 : caractères accentués et spéciaux («ASCII supérieur»)

La partie 0-127 est un standard international et commun à tous les systèmes d'exploitation.

La partie 128-255 diffère entre Mac et PC par exemple, ou entre Windows et MS-DOS (pour ceux qui ont un souvenir de cet ancêtre...), et le codage comporte différentes variétés régionales. Cependant, il y a plus un moins un standard accepté pour l'Europe occidentale (certains codes de cette partie ayant cependant des fonctions particulières selon les logiciels utilisés)

Unicode

Le **codage le plus récent est l'UNICODE**, un codage sur 2 octets (soit $256 \times 256 = 65536$ valeurs possibles) ou davantage au besoin, qui permet ainsi de coder d'innombrable caractères, avec une zone réservée pour chaque alphabet particulier (dont le thaïlandais, le laotien, le cambodgien... et même les milliers d'idéogrammes chinois). C'est le codage qu'il faut utiliser sur l'internet, et vers lequel il est recommandé d'évoluer dans la mesure où les logiciels utilisés le permettent (ce qui n'est pas le cas de tous, notamment en raison du problème de la séparation syllabique).

Types de codage basés sur le clavier US ou européen

Type clavier-clavier

Il s'agit d'un type de codage basé sur la *correspondance entre le clavier américain et le clavier laotien*, permettant de dactylographier sans gestionnaire de clavier. Par exemple, la lettre « koo » (ກູ້) en laotien se tape avec la touche D du clavier américain, car c'est à cet endroit que se trouve cette lettre laotienne. Évidemment, ça ne marche plus si on a un clavier français ou suisse... Les caractères laotiens sont donc codés dans la partie ASCII inférieur, à la place des caractères latins.

Parmi ces polices, on trouve notamment la série *Alice_0*, *Alice_1*, etc. (polices nommées avec un caractère de soulignement avant le numéro!)

Type clavier-phonétique

Il s'agit d'un type de codage basé sur une correspondance approximativement phonétique avec le clavier américain ou français... Par exemple, le « koo » (ກູ້) laotien se tape avec la lettre K du clavier. Cela donne lieu à diverses variantes qui rendent ces polices incompatibles entre elles, et des textes illisibles si l'on a pas exactement la bonne police. Les caractères laotiens sont également codés dans la partie ASCII inférieur, à la place des caractères latins.

L'avantage de ces codages est de pouvoir dactylographier sans gestionnaire de clavier spécial. Mais elles ne permettent pas de faire de vrai traitement de texte, posent des problèmes de compatibilité en raison de la diversité des codages, et sont à éviter pour tout document destiné à être partagé ou publié.

On pourra installer au besoin ces polices pour pouvoir lire des documents qui les utilisent, mais il sera préférable de convertir les documents...

Types de codage standard proposés par Lao Script for Windows (LSWin)

LSWin propose un codage basé sur la logique et l'ordre de l'alphabet laotien, avec un gestionnaire de clavier permettant la dactylographie selon le clavier laotien standard qui prend en charge la **séparation syllabique**. Le codage comporte plusieurs variantes dues à l'évolution de Windows et de MS Word depuis la création du système.

Le gestionnaire de clavier permet également (depuis la version 6) la dactylographie au moyen d'un système de codage romanisé, interprété par le logiciel :

on dactylographie par exemple pour produire *Kian phaa saa laav pen too Nang seu frangl*
ຂໍາພາສາລາວເປັນໄຕໜັງສີຟຣົງ

Ce codage garde les caractères latins non accentués dans la partie ASCII inférieur, et code le laotien dans la partie ASCII supérieur, à la place des caractères accentués et spéciaux. Il permet ainsi de mélanger laotien et anglais sans changer de police. Il comporte un code spécial pour indiquer les espaces invisibles entre les mots ou les syllabes qui permettent de faire du traitement de texte.

Les deux variantes principales sont les suivantes :

LSWin

Ce codage est utilisé en standard sur les ordinateurs du SFE depuis le début, et c'est celui qu'utilisent les personnes ayant installé LSWin depuis ses débuts (avant 2000).

Lao2000

Il s'agit d'une modification du codage d'origine, ce dernier ayant dû être remanié avec l'apparition des version 2000 de Microsoft, car certains caractères ne fonctionnaient plus – ils fonctionnent d'ailleurs de nouveau avec des versions ultérieures!). C'est le codage que utilisé actuellement pour le magazine *Contact*.

Unicode

L'Unicode, reconnu internationalement, est le standard de l'avenir. Le laotien est codé dans la plage qui lui est réservée, et la police comporte tous les caractères européens standards à leur place respective.

LSWin recommande ce codage Unicode, avec les polices analogues de même graphie en version Unicode.

LSWin recommande le codage Unicode, avec les polices suivantes de même graphie en version Unicode. La police la plus standard, *Saysettha Lao*, qui existe dans les différents codages proposés par LSWin. Les polices *Alice* sont aussi disponibles dans ces différents codages. Elles sont nommées Alice1, Alice2, etc., sans le caractère de soulignement, Alice1 2000 et Alice1 Unicode, respectivement.

Cependant, si les logiciels de traitement de texte comme Word et OpenOffice supportent le caractère spécial de séparation syllabique, ce n'est pas encore le cas général dans les logiciels de mise en page, ce qui limite les possibilités.

Recommandation importante

L'usage de polices utilisant l'un des types de codage basés sur le clavier US ou européen est à proscrire pour tous les documents destinés à être partagés.

Les documents utilisant ces polices ne permettent aucune mise en page, en raison de l'absence de code de séparation de mots. Le dactylographe doit alors introduire manuellement un retour de ligne forcé au bout de chaque ligne, et toute correction ultérieure oblige à réviser manuellement les fins de ligne.

Mais **des conversions sont possibles**. LSWin fournit une extension de MS Word qui permet notamment d'effectuer ces conversions avec la plupart des polices. Tout document créé avec un type de police exotique devrait donc être converti dans un des formats proposés par LSWin.

Note sur la séparation des mots et des syllabes

Le problème de la séparation des mots dans une langue qui « colle » les mots comme le laotien n'a malheureusement pas de solution universelle, et c'est une des raisons qui a obligé le logiciel LSWin à modifié le codage utilisé au fil de l'évolution des logiciels et du système d'exploitation de Microsoft.

Lao Script for Windows : <http://www.laoscript.net/>

Tables de codage des caractères latins et de l'alphabet laotien

Alphabet latin, police Verdana, code standard Windows (occidental, ISO-8859-1)

Alphabet laotien, police Alice_5, code Alice_xxx, type clavier-clavier

Exemple de code type clavier-phonétique

À compléter

Codages proposés par LSWin

Alphabet laotien, police Saysettha Lao (code LSWin)

32	! " # \$ % & ' () * + , . /	0 1 2 3 4 5 6 7 8 9 :	; < = > ?
64	@ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z [\] ^ _		
96	' a b c d e f g h i j k l m n o p q r s t u v w x y z { } ~		<small>DEL</small>
128	€ □ , Š • - □ Ž □ □ ‘ ’ “ ” • —		□ ž Ÿ
160	ກ ໂ ຕ ເ ຈ ສ ແ ຍ ດ ຕ ຖ - ບ ປ ຜ ພ ພ ພ ພ ພ	໌ ວ ອ ອ ຕ ຈ	໌ ຊ ຊ
192	ເ ແ ໂ ໃ ໄ ໅ ໆ ຊ ຊ ພ ພ ພ	໌ ຊ	ບ
224	ອ ແ ໂ ໃ ໄ ແ ດ ນ ຖ ດ ດ ດ	ີ ດ	ີ

Alphabet laotien, police Saysettha 2000 (même graphie, code Lao2000)

32	! " # \$ % & ' () * + , - . /	0 1 2 3 4 5 6 7 8 9 :	; < = > ?
64	@ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z [\] ^ _		
96	' a b c d e f g h i j k l m n o p q r s t u v w x y z { } ~		<small>DEL</small>
128	€ □ , ເ „ … ແ %oo □ Ž □ □ ‘ ’ “ ” • — ™ ແ ໂ ແ ແ		໌ ຊ ຊ
160	¢ £ ₣ ¥ ! © « - ® ° ໄ . » ¼ ½ ¾		
192	ກ ໂ ຕ ເ ຈ ສ ແ ຍ ດ ຕ ທ ບ ປ ຜ ພ ພ ພ ພ	໌ ວ ອ ອ ຕ ຈ	
224	໌ ຊ ແ ໂ ໃ ໄ ແ ດ ນ ຖ ດ ດ ດ	ີ ດ	ີ

Unicode

Pour Unicode, le laotien est codé dans la page du laotien, en dehors de la table ASCII, ce qui permet de mélanger le laotien et le français (et même d'autres langues) sans changer de police.

Voir le tableau à cette adresse : <http://unicode.org/fr/charts/PDF/U0E80.pdf>