

How to Type Khmer Unicode

If you have used the old-style Khmer fonts (Lemon, Battambang, ABC, etc.) before, it will take you very little time to get used to the new fonts (Unicode fonts). In only one or two days you will be typing faster than before.

Unicode is not just one more font. Unicode is the new International standard for Khmer. Computers all around the world will understand Khmer language when typed using Unicode fonts. Unicode is also the new standard of the Cambodian Government. In a few years all the old fonts will be forgotten, and all texts will be in Unicode.

All Unicode fonts have the same encoding, meaning that once a text is typed in a Unicode font, it can be changed to any other Unicode font, and the text will remain exactly the same, something that did not happen with the old fonts.

One of the most important differences between the old of typing and Unicode is that when using the new fonts, there is a key in the keyboard for each vowel, including េៀ, េ្នី, េ្នា, េ្នា, េ្នី, េ្នំ, េ្នះ, etc. In Unicode you type the letters as they are in Khmer, not some parts of them first and then the rest. And most important, you type them in the order that they are spelled, not from left to right, as you would write them by hand. As an example, if you wanted to type the word លើ with the old fonts, you had to write

េ្នំ + ល + េ្នី,

With the new fonts you would type

ល + េ្នី

You can see that not only the keys that you type are different; you also type first the consonant and then the vowel, as if you were spelling the word (LO... SRA OE).

Another big novelty in the new way of typing is that the coeng consonants are not in the keyboard, In order to write them, you need to use two different keys: a key to indicate the following consonant is a coeng consonant, and then the normal key for that consonant. For example, to write the word អ៊ី, you need to type:

្រី + ្រ + ឃី + ្រ

This is done this way in order to type faster, and so that you do not need to remember the situation in the keyboard of all the coeng consonants.

But these are just some examples to introduce you to the new way of typing. We will now start from the beginning and go through all the details one by one, starting by looking at the Khmer letters themselves.

Khmer Consonants, Vowels and signs

Consonants:

ក	ខ	គ	ឃ	ង			
ច	ឆ	ជ	ឈ	ញ			
ដ	ប	ឧ	ធរ	ណ			
ត	ថ	ទ	ធ	ណ			
ប	ផ	ព	ភ	ម			
យ	រ	ល	វ	ស	ហ	ឡ	អ

Independent Vowels

Modern Khmer has 14 independent vowels:

ឥ	ឡ	ឧ	ឧ	ឌី	ឡ	ឬ
ឬ	ឡ	ឡ	ង	ឌី	ឧ	ឌី

Dependent Vowels

The Chuon Nat dictionary, a classic reference for Khmer, lists 20 vowels:

ា	ិ	ី	ី	ី	ុ	ុ
ូ	ើ	ៀ	ៀ	េ	ៃ	ៃ
ៅ	ៅ	ុំ	ំ	ាំ	ះ	

Nevertheless, the modern Khmer system taught in schools lists three more vowels, raising the number to 23. These vowels – also included in the keyboard - are:

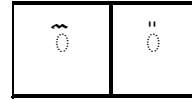
ុះ	េះ	ៅះ
----	----	----

There are other combinations that – following the same criteria – (combining a vowel with the REAHMUK mark) could also be considered as vowels, but they are not considered as such by any source, so they are not included in the keyboard. They have to be written with two keystrokes. They are:

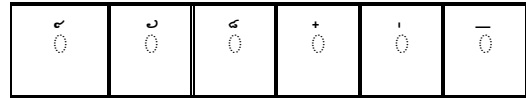
ើះ	ីះ	ៃះ
----	----	----

Consonant shifters

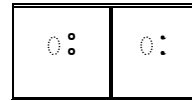
Triisap and Muusikatoan. When followed by a high vowel, they change their shape to that of vowel SRA U.

**Signs placed above**

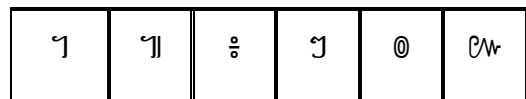
Toandakhiat, Samyok sannya, Ashda, Kakabat, Bantoc and Viriam.

**Robat****Signs placed behind the consonant**

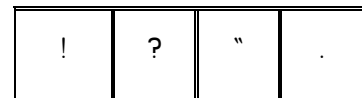
Reahmuk and Yuukaleapintu.

**Khmer punctuation signs**

Khan, Bariyoosan, Camnuc Pi Kuu, Lek Too, Phnaek Muan and Koomuut

**Latin punctuation signs used in Khmer**

Khan, Bariyoosan, Camnuc Pi Kuu, Lek Too, Phnaek Muan and Koomuut



Khmer numbers

១	២	៣	៤	៥
៦	៧	៨	៩	០

Riel currency symbol

៛

Khmer divination numbers

៦	៧	៨	៩	០
១	២	៣	៤	៥

Lunar dates

០ ៨	១ ១	២ ១	៣ ១	៤ ១	៥ ១	៦ ១
៧ ១	៨ ១	៩ ១	១០ ១	១១ ១	១២ ១	១៣ ១
១៤ ១	១៥ ១	៨ ៨	១ ១	២ ១	៣ ១	៤ ១
៥ ១	៦ ១	៧ ១	៨ ១	៩ ១	១០ ១	១១ ១
១២ ១	១៣ ១	១៤ ១	១៥ ១			

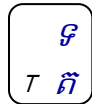
Other signs used for old Khmer, or for Sanskrit or Pali transliteration

Letter SHA, letter SSO, sign Avakrahasannya, independent vowel QUK, Sign Atthacan and sign Bathamasat

ឝ	ឞ	ៗ	ឝ	˘	◌̣
---	---	---	---	---	----

The Keyboard

The computer in which you want to type in Khmer script needs to know that what you want to type are Khmer letters, so that when you press the key:



the computer knows that it has to print the letter ្គ, and not the letter T.

You tell the computer that you want to write in Khmer by “selecting the Khmer Keyboard”. You can see what keyboard is selected in your computer by looking at the bottom-right corner of your screen, which looks something like:



Please pay special attention to the letters EN. EN is short for English. If you see this letter, it means that the computer is using an English keyboard, so for every key that you type, you will always get English letters.

If, nevertheless, you see this:



The letters CA mean that your computer is using a Khmer keyboard, and that therefore any key that you hit will produce Khmer characters.

You can easily change between the English and the Khmer Keyboards, by either.

- **Pressing the Shift and the Alt keys down at the same time.** Every time that you press them there will be a change of keyboard. If the present keyboard is the English one (EN), it will change to Khmer (CA), and if the current one is the Khmer keyboard, it will change to English.
- **Clicking on the letters EN with your mouse.** When you do this, a small window with all the possible options will open. In this window you should click on CA (ignore the fact that besides CA it says “Catalan”).

To test all the examples in this document, you need to use the Khmer keyboard; the bottom-left corner of your screen should display the letters CA.



Typing in Khmer Unicode

When using Unicode, typing is done in spelling order, not in left-to-right writing order. The computer itself takes care of putting the letters in the right order. As we mentioned before, a simple word like:

លើ

is typed:

ល + ើ

and not

េ + ល + ើ

This order must be followed for all words. The general rule is that syllables must be written in the order:

Consonant + Coeng Consonant(s) + Vowel + sign(s)

Consonants and independent vowels are used in the same ways in Khmer orthography, so when we say consonant - in this document - it will generally mean “consonant or independent vowel”.

Vowels are typed always after the consonant, independently of them being placed before, after, above or below the consonant. The computer will place them in the right place.

Vowels that have two parts (such as េ) are also typed after the consonant:

ល + ើ → លើ

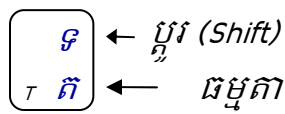
គ + េ → គេ

ន + ៅ → នៅ

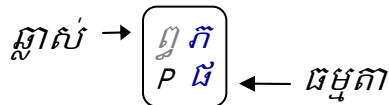
ល + ើ → លើ

If you type letter in an incorrect order, the computer display a dotted circle to indicate that the construction is incorrect. For example, if you type: រ + ល + ី the computer will display រលី

The consonants and vowels placed in the lower part of the key are typed by just pressing the corresponding key. Consonants and vowels placed in the higher part of the right side of the key are typed by holding the SHIFT key pressed and pressing that key.

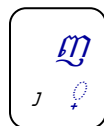


Some letters that are not often used (some independent vowels), and which are printed in the top-left side of a key in grey color, need to be typed by pressing the Alt Gr key (right-side Alt key, ឆ្ល័ស).



If a syllable contains any coeng consonants, they have to be typed after the main consonant, but before the vowels, if there are any.

You might have noticed that the coeng consonants are not in the keyboard, as was the case before with keyboards for Limon, ABC or other fonts. This is because in Unicode the same key is used to type the normal consonant and the coeng consonant. In order to signal that the next consonant must be a coeng consonant, the following key must be typed before the consonant:



For example:

ច + តិ + ្ក + តិ → ចត្តិ
 ស + ្ក + រ + ី → ស្រី
 ស + ្ក + រ + ៃ → ស្រៃ

For syllables that have two coeng consonants, you must type both coeng consonants in spelling order, each one of them preceded by the ្ក sign:

ស + ្ក + តិ + ្ក + រ + ី → ស្រីត្រី

Make sure that you **do not type coeng RO before another coeng consonant**, always follow spelling order, and **never type a coeng consonant after a vowel**.

Special cases for consonants

There are some cases in which letters can be represented with more than one shape, depending on others letters around them. This is the case of letter ប when followed by ិ. In the old fonts it was necessary to have a special letter for this, but in Unicode you must type the normal key for ប. When you type ិ the computer will find the right shape of letter ប.

ប + ិ → បិ

Something similar happens with letter ញ. You do not have to care about the right shape of ញ or ្ណ. The computer will always find the correct one. You only need to make sure that you type the right letters.

ទ + ិ + ញ → ទិញ
 ស + ញ + ្ក + ញ + ិ → សញ្ញា
 ខ + ្ក + ញ + ្ក → ខ្ញុំ
 ប + ញ + ្ក + ច + ៃ + ញ → បញ្ចេញ

Never type ព + ិ instead of ញ.

Special cases for vowels

There are some vowels, that, in spite of being in the keyboard, correspond to two different characters (they do not have their own characters in Unicode). These vowels are:

ាំ ៉ំ ុំ ៃ៖ and ៃៗ៖

You can type these vowels directly with their keyboard key, but if you later want to erase them; you will have to press the ← key twice.

The opposite happens with vowels that have two parts (split vowels):

ើ ើៗ ើៗ ៃៗ and ៃៗ

They also type by pressing only one key, but they are erased by pressing ← only once.

Other vowels that are not in the keyboard require typing two keys. For example:

ក្រំ + ៃៗ + ុំ → ក្រុំ

Consonant-shifters

Sometimes, a character called a consonant-shifter is used to change the sound of a vowel that accompanies a consonant. There are two of these characters:

Triisap (៉) is used to change the sound vowels that accompany consonants of the first series to sound as if they accompanied consonants of the second series. It can only be used with consonants of the first series that do not have a second series equivalent. These consonants are:

ស ហ អ and ឃ

Muusikatoan (៉) is used to change the sound vowels that accompany consonants of the second series to sound as if they accompanied consonants of the first series. It can only be used with consonants of the second series that do not have a first series equivalent. These consonants are:

ង ញ ម យ វ and រ

As an exception, ៉ is also used with the letter ប to change its sound from a “b” sound to a “p” sound.

The consonant-shifter is always typed after the consonant and any coeng consonants, but before the vowel. The order of components, including the consonant-shifter, will be:

Consonant + Coeng Consonant(s) + Consonant-shifter + Vowel + sign(s)

Examples:

យ + ៉ + ា + ក → យ៉ាក
 ប + ៉ + ោ + ង + ប + ៉ + ោ៖ → ប៉េងោ៖
 ក + ុ + ុ + ម + ហ + ៉ + ុ + ន → ក្រុមហ៊ុន
 ប + ៉ + ុ + ត → ប៉ុត
 ស + ុ + រ + ៉ + ុ + ប → ស្រុប
 ម + ុ + ហ + ៉ + ៃ + ត → ម៉ៃត

When the consonant-shifter is either followed by a vowel that needs to be placed above the consonant or by ាំ, the consonant shifter changes its position automatically to be placed under the main consonant, and also changes its shape to the same shape of the vowel ុ (same for ៉ and for ៉). **You do not have to type ុ**. If the lower-form shape is necessary, the computer itself will display it that way. Some examples:

ស + ៉ → ស៉

$$\begin{aligned}
\text{ស} + \tilde{\text{o}} + \text{័} &\rightarrow \text{ស័} \\
\text{ប} + \text{័} + \text{័} &\rightarrow \text{ប័} \\
\text{ប} + \text{័} + \text{័} + \text{ស} + \tilde{\text{o}} + \text{័} &\rightarrow \text{ប័ស័}
\end{aligned}$$

If you type ័ instead of $\tilde{\text{o}}$ or ័ , it will not work, as you cannot type two vowels one after the other in Khmer:

$$\text{ស} + \text{័} + \text{័} \rightarrow \text{ស័័}$$

There are three exceptions to the rule in which the consonant-shifter does not change to the lower position, in spite of having something above the consonant. You do not have to really worry about them, because they are automatically handled by the computer, but you should look at them, so when they appear you are not surprised or think that something is not working correctly.

- It does not change when the vowel is ័ , because the combination of ័ with the lower shape of the consonant shifter could be confused with the vowel ័ . The same happens for all signs placed above the consonant. If they are ever combined with a consonant-shifter, that consonant-shifter will be located above the consonant.

$$\text{អ} + \tilde{\text{o}} + \text{័} \rightarrow \text{អ័}$$

- It cannot change when the letter ប is combined with $\tilde{\text{o}}$, because the reader would not know if the lower form of the consonant shifter represents $\tilde{\text{o}}$ or ័ (therefore if is the sound of the vowel that must change or is the sound of the consonant should be “p” instead of “b”).

$$\text{ប} + \tilde{\text{o}} + \text{័} + \text{ក} \rightarrow \text{ប័ក}$$

- When there is a main consonant and a coeng consonant, and either of them is ឃ . This rule is followed consistently all over the Chuon Nat dictionary. This is

probably due to the fact that the other consonant in the syllable might be of the second series, and if the consonant shifter is placed down, it is not clear if it is a ័ that applies to ហ៊ or a ័ that applies to the other consonant.

អ + ្ក + ហ៊ + ័ + ិ + ៈ → អ័្កិៈ

Signs

There are three types of signs in Khmer that may form part of a syllable:

- **Robat** (័). Robat is a sign borrowed from Indian languages that is phonetically equivalent to placing a RO sound before the consonant above which it is placed. It is treated differently from other signs. ័ must always be typed after the consonant above which it will be placed, even before any vowels are typed. A syllable that has ័ cannot have any coeng consonants, nor vowels or signs that need to be placed above the main consonant. When Robat is used, the syllable can only be constructed as:

Consonant + Robat {+ Vowel} {+sign}

As in:

ព + ណ + ័ → ពណ័

ស + គ + ័ + ៈ → សគ័ៈ

ក + ប + ័ + ្ក + រ → កប័្ករ

ក + ា + ប + ័ + ា + ស → កាប័ាស

ស + ព + ័ + េ + ជ + ្ក + ញ → សព័េជ្កញ

- Other signs placed above the main consonant (័ ័ ័ ័ ័ ័). Toandakhiat (័), Samyok Sannya (័), Ahsda (័), Kakabat (័), Viriam (័) and Bantoc (័), are placed above the main consonant. Normally they do not combine with vowels, but in the cases they do, they should always be typed after the vowel, and

before any signs placed behind the syllable. When they combine with a vowel, in most case the vowel is ឺ (not to be confused with the consonant-shifters ឺ or ឺ). If we include all this in our formula, we get:

Consonant + coengs + consonant-shifter + vowel + Above signs + After signs

$\text{ល} + \text{ក} + \text{◌} \rightarrow \text{លក}$
 $\text{ទ} + \text{◌} + \text{ព} + \text{◌} + \text{រ} \rightarrow \text{ទ័ព័រ}$
 $\text{ណ} + \text{◌} + \text{◌} \rightarrow \text{ណ៎}$
 $\text{ន} + \text{◌} + \text{◌} + \text{◌} \rightarrow \text{ន៎}$

There are two exceptions, which are automatically handled by the computer, but please pay attention to them:

- Sometimes a high vowel might combine with Kakabat as in:

$\text{អ} + \text{◌} + \text{◌} + \text{◌} \rightarrow \text{អ៎}$

- The ឺ + ឺ combination, used in many words. Some of the old fonts had a special character for this. In Unicode the two characters are correctly ligated when you type them separately, as in:

$\text{ព} + \text{◌} + \text{◌} + \text{◌} + \text{◌} \rightarrow \text{ព័}$

Technical note: As a short technical note, and in spite of us calling it vowel here, to simplify the reading, Unicode considers ឺ as a sign, and not as a vowel. The characters ឺ and ឺ need to be formed by combining the vowels ឺ and ឺ with the sign ឺ (sign Nikahit). Both Khmer vowels (ឺ and ឺ) exist in the keyboard, for compatibility with Khmer culture, but what the keyboard really does is to generate the characters ឺ or ឺ and then the character ឺ. The same happens with the vowels ឺ, ឺ and ឺ. They are created automatically by the keyboard, but we have to understand that when we press one of these vowels, the keyboard is first doing a vowel (ឺ, ឺ or ឺ) and then the sign Reahmuk (ឺ).

- **Sign placed after the consonant which are part of the syllable and affect the sound (ឺ and ឺ).** Reahmuk (ឺ) and Yuukaleapintu (ឺ) are always placed as the last

element of the syllable. They should not be confused with the punctuation mark Camnuc pii Kuu (៖).

ក្រី + ក្រី + ៖ → ក្រី៖

Punctuation marks (ៗ ៗ ៖ ៗ ៖ ៗ ! . " ?)

Khmer script has its own punctuation marks: sign Khan (ៗ), sign Bariyoosan (ៗ), sign Camnuc pii Kuu (៖), sign Leik Too (ៗ), sign Phnaek Muan (៖) and sign KooMuut (ៗ).

In addition to these, some punctuation marks are borrowed from the Latin script, such as the exclamation sign (!), period (.), double quotes (“ ”) and interrogation mark (?). The period is only used for abbreviations. With the exception of ៗ, Khmer punctuation marks are usually separated from the preceding and following words by a space. Latin punctuation marks are usually not separated from the word that precedes them, except for the quotation marks, which is attached to the word that follows it when opening a quotation and attached to the word that precedes it when closing the quotation.

In Khmer, the space (ៗ) is also considered as a punctuation mark, and should therefore be used only to mark a pause in the speech, not to separate words. In order to type a space, you need to hold the SHIFT key down and at the same time press the spacebar.



About ligatures

In Unicode fonts, when a consonant is followed by ៗ, it changes slightly its shape for aesthetic reasons. IF the ៗ is erased, the consonant goes back to its original form. The user does not normally know that ligatures happen; he only knows that his text looks nice. Please pay attention to this:

ក្រី ៗ ក្រី

You can see how the shape of the line above រ៉ី has changed, just to make the two letters look nicer.

Another traditional ligature of Khmer happens when a Moul letter UO is accompanied by the vowel SRA I. The ligature is written as:


 Instead of
 

Zero Width Space (ZWSP).

In Khmer words are written one after the other, unless punctuation rules specify that they must be separated. Word processing programs, Internet browsers and other programs that format text need to know where they can cut a sentence in order to start a new line. In western languages that use the space to separate the words, this is not a problem, but in Khmer language all these programs do not know how to recognize the end of a word, and therefore a place where the line can be broken so that a new line can be started.

In most newspapers in Cambodia, the line breaking is done by hand. The person writing the text calculates if the next word will fit in the line or not, and he or she will go to the line by hand. This system works for formatting texts that will never be changed.

Web pages cannot use this system; the format of the page depends on the size of the browsing program window in the user's computer, and on the resolution of the screen. Each case is different, so the program itself needs to do the line breaking, a static line-breaking system does not work. The same happens when working on good formats for text. Using a program flexibly requires the program to be able to break the text between words.

The solution to this problem is to insert between words a character that – like a space – has no picture, but, unlike the space, does not separate the words graphically, a gap between the words that cannot be seen. A gap that programs can use to know where to break the lines.

This character is called Zero Width Space, and, it is the most used character in a text (out of each 100 key presses while typing Khmer, 16 are on the Zero width Space), so this highly used character is placed in the keyboard in a very easy to access location: the space bar. Every time you press the space bar you insert a separation gap.

If you press the spacebar (without pressing any other key), the character that will be added to the text is a ZWSP, not a visible space.

If you wish to type a space, you must hold down the SHIFT key and then press the spacebar.

Typing practice

If you have read down to here, you should by now already understand the mechanics of typing Khmer using Unicode fonts. It is time that you put everything in practice, by typing the following text:

Advanced typing

Khmer has other characters that are very seldom used. These characters - used for old Khmer or for transliteration of Pali and Sanskrit - can be accessed using the Khmer keyboard, but they are not printed in the keyboard. The characters are:

Old Khmer, Sanskrit and Pali

Alt Gr + B → ឃ

Alt Gr + K → ឃ

Alt Gr + T → ឃ

Alt Gr + Q → ្ក

Alt Gr + W → ្ក

Shift + Alt Gr + U → ្ក

Shift + Alt Gr + I → ្ក

Shift + Alt Gr + O → ្ក

Shift + Alt Gr + P → ្ក

Shift + Alt Gr + [→ ្ក

Shift + Alt Gr +] → ្ក

Shift + Alt Gr + A → ្ក

Shift + Alt Gr + S → ្ក

Shift + Alt Gr + D → ្ក

Shift + Alt Gr + F → ្ក

Shift + Alt Gr + G → ្ក

Shift + Alt Gr + H → ្ក

Shift + Alt Gr + J → ្ក

Shift + Alt Gr + K → ្ក

Shift + Alt Gr + L → ្ក

Shift + Alt Gr + ; → ្ក

Shift + Alt Gr + ' → ្ក

Shift + Alt Gr + Z → ្ក

Shift + Alt Gr + X → ្ក

Shift + Alt Gr + C → ្ក

Shift + Alt Gr + V → ្ក

Shift + Alt Gr + B → ្ក

Shift + Alt Gr + N → ្ក

Shift + Alt Gr + M → ្ក

Shift + Alt Gr + , → ្ក

Divination signs

Shift + Alt Gr + 1 → ្ក

Shift + Alt Gr + 2 → ្ក

Shift + Alt Gr + 3 → ្ក

Shift + Alt Gr + 4 → ្ក

Shift + Alt Gr + 5 → ្ក

Shift + Alt Gr + 6 → ្ក

Shift + Alt Gr + 7 → ្ក

Shift + Alt Gr + 8 → ្ក

Shift + Alt Gr + 9 → ្ក

Shift + Alt Gr + 0 → ្ក

Lunar dates:

Shift + Alt Gr + Q → ្ក

Shift + Alt Gr + W → ្ក

Shift + Alt Gr + E → ្ក

Shift + Alt Gr + R → ្ក

Shift + Alt Gr + T → ្ក

Shift + Alt Gr + Y → ្ក

Shift + Alt Gr + . → ្ក